PROTEOMIC PROFILING OF METASTATIC MATRISOME REVEALS CITRULLINATION AS A MARKER OF COLORECTAL LIVER METASTASIS

Mr. Arseniy Yuzhalin

University college

A thesis submitted to the Medical Sciences Division of the University of Oxford in partial fulfilment the requirements for the degree of Doctor of Philosophy

CRUK/MRC Oxford Institute for Radiation Oncology

Department of Oncology

University of Oxford

Hilary 2018
Word count: 33451

ABSTRACT

PROTEOMIC PROFILING OF METASTATIC MATRISOME REVEALS CITRULLINATION AS A MARKER OF COLORECTAL LIVER METASTASIS

Submitted for a degree of Doctor of Philosophy
Mr. Arseniy Yuzhalin, University College, Hilary Term 2018

Colorectal cancer is one of the most frequently occurring malignancies and a major cause of cancer death. Distant metastases in this disease most commonly develop in the liver and are often untreatable. Here, we use proteomics to characterise, qualitatively and quantitatively, extracellular matrix (ECM) from colorectal cancer liver metastases. We show that citrullination of the ECM by cancer cell derived peptidyl arginine deiminase 4 (PAD4) is important for the growth of liver metastases. Citrullination of proteins, a post-translational conversion of arginine residues to citrulline, is well recognised in rheumatoid arthritis, but largely undocumented in cancer. PAD4, a key enzyme responsible for catalysing citrullination, is produced by metastatic colorectal cancer cells and found at higher levels in human liver metastases than in normal liver. Functional significance for citrullination in metastatic growth was evident in murine models where inhibition of citrullination, either globally by pharmacologic inhibition of PADs or specifically in colorectal cancer cells by PAD4 knockdown reduced liver metastatic burden by 3- to 4-fold ($P < 0.05$). Additionally, citrullination of key extracellular matrix (ECM) component collagen type I *in vitro* led to greater adhesion and 25-30% decreased migration of colorectal cancer cells ($P < 0.05$) along with increased expression of characteristic epithelial markers, indicating a role for citrullination in promoting mesenchymal-to-epithelial transition (MET). Overall, our study revealed PAD4-dependent citrullination of the ECM altering mesenchymal-epithelial plasticity in colorectal cancer cells and the progression of liver metastasis. These data indicate that inhibition of citrullination could be exploited to potentially prevent the development of liver metastases in colorectal cancer.

ACKNOWLEDGEMENTS

First, I would like to acknowledge my group leader Professor Ruth Muschel for enabling me to accomplish this project. I would like to also thank my lab members Esther Lim, Alex Gordon-Weeks and Keaton Jones for their immense help with experiments.

Additionally, I'm greatly indebted to my mother Olga Kozlova who helped in typing, formatting and referencing this thesis.

Finally, this work would never have happened without funding provided by Cancer Research UK and Medical Research Council.

SPECIFIC ACKNOWLEDGEMENTS

Figure 1 - experiments were performed with help from Esther Lim

Figure 5 - experiments were performed with help from Rebecca Konietzny and Benedikt Kessler

Figure 6 - experiments were performed with help from Rebecca Konietzny and Benedikt Kessler

Figure 7 - experiments were performed with help from Rebecca Konietzny and Benedikt Kessler

Figure 9 - experiments were performed with help from Rebecca Konietzny and Benedikt Kessler

Figure 10 - experiments were performed with help from Rebecca Konietzny and Benedikt Kessler

Figure 12 - experiments were performed with help from Alex Gordon-Weeks

Figure 13 - experiments were performed with help from Alex Gordon-Weeks

Figure 14 - experiments were performed with help from Alex Gordon-Weeks

Figure 16 - experiments were performed with help from Alex Gordon-Weeks and Andrew Worth

Figure 18 - experiments were performed with help from Alex Gordon-Weeks

Figure 21 - experiments were performed with help from Rebecca Konietzny and Benedikt Kessler

Figure 23 - experiments were performed with help from Rebecca Konietzny and Benedikt Kessler

Figure 28 - experiments were performed with help from Rebecca Konietzny and Benedikt Kessler

Figure 29 - experiments were performed with help from Alex Gordon-Weeks

Figure 32 - experiments were performed with help from Eric O`Neill, Nikola Vlahov and Maria Tognoli

Figure 33 - experiments were performed with help from Eric O`Neill, Nikola Vlahov and Maria Tognoli

Figure 35 - experiments were performed with help from Rebecca Konietzny and Benedikt Kessler

Figure 37 - experiments were performed with help from Graham Brown and Bostjan Markelc

Figure 40 - experiments were performed with help from Alex Gordon-Weeks and Keaton Jones

Figure 43 - experiments were performed with help from Alex Gordon-Weeks and Keaton Jones

Figure 44 - experiments were performed with help from Alex Gordon-Weeks, Ruth Muschel and Keaton Jones

Figure 45 - experiments were performed with help from Abul Siddiky

LIST OF ABBREVIATIONS

ACPA - anti-citrullinated protein antibodies

ANOVA - analysis of variance

APC - adenomatous polyposis coli

ATCC - American Type Culture Collection

BMI - body mass index

BMP-7 - bone marrow protein 7

Bz-ADMA - benzoyl-N^G,N^G-dimethyl-Arg

CCP2 - cyclic citrullinated peptide

CMV - cytomegalovirus

CT - computed tomography

CTGF - connective tissue growth factor

DAPI - 4',6-diamidino-2-phenylindole

DMEM - Dulbecco's Modified Eagle Medium

DMH - dimethyhydrazine

DMSO - dimethyl sulfoxide

DTT - dithiothreitol

ECM - extracellular matrix

EDTA - ethylenediaminetetraacetic acid

EGF - epidermal growth factor

ELISA - enzyme-linked immunosorbent assay

EMT - epithelial-to-mesenchymal transition

ERK - extracellular signal-regulated kinase

FACS - fluorescence activated cell sorting

FAK - focal adhesion kinase

FAP - familial adenomatous polyposis

FBS - foetal bovine serum

FDR - false discovery rate

FGF2 - fibroblast growth factor 2

GO - gene ontology

GFAP - glial fibrillary acidic protein

GFP - green fluorescent protein

HSEC - hepatic sinusoidal endothelial cells

HEPES - N-2- hydroxyethylpiperazine-N- 2-ethane sulfonic acid

HLA - human leukocyte antigen

HLF - hepatic leukaemia factor

HNPCC - hereditary nonpolyposis colorectal cancer

HPRT - hypoxanthine phosphoribosyltransferase

ICAM - intercellular adhesion molecule 1

IHC - immunohistochemistry

IL-6 - interleukin 6

ING4 - inhibitor of growth protein 4

JPS - juvenile polyposis syndrome

LC-MS/MS - liquid chromatography–mass spectrometry

MAM - methylazoxymethanol glycoside

MBP - myelin basic protein

MEF - mouse embryonic fibroblasts

MET - mesenchymal–epithelial transition

MMP - matrix metalloprotease

MRI - magnetic resonance imaging

mRNA - messenger RNA

miRNA - microRNA

MS - multiple sclerosis

mTOR - mammalian target of rapamycin

OCT - optimal cutting temperature

PAD - peptidyl arginine deiminase

PBS - phosphate-buffered saline

PDGF - platelet-derived growth factor

PFA – paraformaldehyde

PhIP - 2-amino-1-methyl-6-phenylimidazo[4,5-b]-pyridine

PJS - Peutz-Jeghers syndrome

PCR - polymerase chain reaction

RA - rheumatoid arthritis

RF - rheumatoid factor

RPMI - Roswell Park Memorial Institute

RT - room temperature

SCID - severe combined immunodeficiency

SDS - sodium dodecyl sulphate

SDS-PAGE - sodium dodecyl sulphate polyacrylamide gel electrophoresis

SLE - systemic lupus erythematosus

STAT3 - signal transducer and activator of transcription 3

TCGA - The Cancer Genome Atlas

TGF - transforming growth factor β

TBS - tris-based saline

tPA - tissue-type plasminogen activator

UK - United Kingdom

VCAM - vascular cell adhesion molecule 1

VEGF - vascular endothelial growth factor

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SUPPLEMENTARY TABLES

CHAPTER 1. INTRODUCTION

1.1. Colorectal cancer

With more than 1 million of new cases per year worldwide, colorectal cancer represents a global public health problem[1]. Epidemiological findings demonstrate that colorectal cancer is the third and second most commonly diagnosed malignancy in Europe in males and females, respectively[2]. In 2012, 447000 new diagnosed cases and 215000 deaths from this disease were recorded in Europe[3]. In the UK in particular, 40000 new cases are registered each year. Notably, there is a large variation in ethnical and geographical distribution of incidence of colorectal cancer. In westernised diet regions of the world, such as North America, Europe and Australia, the incidence of this disease is 10-fold higher in comparison with Africa, Asia and South America. Similarly, age-standardised rates for Caucasian individuals are almost 2-fold higher as compared to Black or Asian patients in both males and females[2]. The explanation for this variation can be environmental factors such as dietary habits and the lifestyle, which are markedly different in these parts of the world[3]. The westernised diet is characterised by high consumption of processed meat and sugar, whilst the intake of fibre is relatively low[4–6]. All the above mentioned parameters are long-term risk factors for colorectal cancer, which is consistent with relatively late median age of colorectal cancer diagnosis (62 and 72 for males and females, respectively)[7]. More than 90% of colorectal cancer patients get diagnosed at the age over 50[3], and it has been shown that it takes 17 years to for a benign polyp to develop into colorectal cancer, making this malignancy predominantly a disease of the elderly[8]. Another explanation of the observed differences in cancer incidence could be a socio-economical variation and low physical activity, resulting in higher body mass index (BMI) difference, which also correlates with increased risk of cancer[9,10].

It is recognised that besides environmental factors, hereditary component is also important in colorectal cancer. Approximately 2-5% of all colorectal cancers arise due to highly penetrant inherited mutations leading to the development of Lynch syndrome (also known as hereditary nonpolyposis colorectal cancer [HNPCC]), familial adenomatous polyposis (FAP), MUTYH-associated polyposis (MAP), Peutz-Jeghers syndrome (PJS) and juvenile polyposis syndrome (JPS)[11]. These conditions are autosomal dominant disorders emerging from DNA repair point mutations leading to polyposis and subsequent colon cancer[12,13]. Also, genetic variability resulting from single nucleotide polymorphisms (SNPs) go some way to explaining differences in colon cancer incidence between different ethnic groups[14].

Emerging evidence suggests that composition of gut microbiota can significantly influence the onset and progression of inflammatory bowel disease[15], irritable bowel syndrome[16], celiac disease[17] and colorectal cancer[18]. Major phyla constituting the gut flora include *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, *Verrucomicrobia* and *Actinobacteria*[19], and their alteration can disrupt assimilation of certain nutrients and modulate host immune response or metabolism. Some enteric flora may initiate colorectal cancer by inducing a long-term inflammation through a release of bacterial toxins[20]. In murine models of chemically-induced colon carcinogenesis, administration of probiotics containing *Lactobacillus* protected the colon against oxidative stress[21] and reduced the number of colonic lesions[22]. Diversity of gut microbiome was recently linked to a better tumour response to anti-PD-1 immunotherapy[23]. In colorectal cancer patients, the diversity of intestinal microbiota was significantly lower than that in the healthy individuals[24]. Despite these intriguing findings, further basic research is needed to

elucidate specifically which bacterial phyla and species influence colorectal cancer development.

Management of colorectal cancer includes surgical resection, radiotherapy, and chemotherapy, individually or in combination[25]. The choice of treatment modality is dependent on the location of a tumour and its stage, grade and individual demographic factors such as physical status and comorbidity, ultimately being a decision between the clinician and patient[26].

Metastasis is the predominant cause of mortality in colorectal cancer patients. The liver is the most frequent site for metastasis, whilst metastases to the lymph nodes, brain, peritoneum, bones, or lung are less common[27]. Population-based epidemiological studies from France[28] and Germany[29] revealed that approximately 25–30% of patients with colorectal cancer develop liver metastases during the course of disease. About 25% of colorectal cancer patients have liver metastases diagnosed at the same time as primary lesions (synchronous metastases), whilst 50% develop liver metastases at some subsequent point post primary diagnosis (metachronous metastases)[30]. Patients with colorectal cancer liver metastasis generally have a poor prognosis[28]. Surgical ablation is the only therapy capable of providing long-term cure for patients with liver metastasis. However, only 20% of patients with hepatic metastases are eligible for surgery, as surgical resection is impossible if the metastatic burden is too large, or the remaining liver part is insufficient to maintain normal functioning or if metastases are close to vital structures such as the hepatic vasculature or biliary tree[30–32]. To increase the number of candidates for surgery whose tumours are considered unresectable, pre-emptive measures such as portal vein embolisation, neoadjuvant chemotherapy or staged hepatectomy are

used to render the liver amenable to curative surgery. Another option, downstaging of inoperable hepatic metastases with different chemotherapeutics (e.g. 5-fluorouracil, irinotecan, oxaliplatin) may enable up to 33% of patients become eligible for surgical ablation. However, the use of chemotherapy drugs can lead to disappearing hepatic metastases in around 40% of cases, whilst the majority of them will again develop macroscopic metastases in future[33,34]. Methods for thermal ablation such as cryotherapy, microwave and radiofrequency ablation or intrahepatic ethanol injection[35–37] are also used to treat liver metastases, however this is primarily in the palliative setting as the long-term cancer-specific survival for these modalities are less well reported.

Patients undergoing liver metastasis resection can expect 5-year overall survival of 40-50%, with up to 10% alive at 10 years[30]. Thus, current management of hepatic metastases from colorectal cancer could be considered ineffective and there is room for significant improvement. Advances in the understanding of the biology of metastatic progression and identification of new therapeutic strategies targeting metastatic disease would therefore be of significant benefit to patient survival.

1.2. Liver anatomy and physiology

Embryologically, the liver arises from the ventral foregut endoderm from the third week of gestation[38,39]. The hepatic primordium consists of cellular cords forming sinusoids, which are supplied with blood from the portal and umbilical veins; upon birth, the latter collapses and becomes obliterated, forming the ligamentum teres[38,39]. During prenatal development, capillary networks gradually fuse, evolving into the left, middle, and right hepatic veins which drain the liver into the inferior vena cava[38,39].

The liver achieves its mature architecture by 15 years of age and represents the largest gland in the human body, normally weighing around 1.5 kg or 2.5% of the total body weight[38]. It occupies the right and partially the left upper quadrant of the abdomen, being enveloped by a thick connective tissue layer named Glisson's capsule, which, in turn, is covered by a thin double-layered visceral peritoneum. The liver is in the indirect contact with the diaphragm at its posteriosuperior surface in a small triangular bare area[39]. The liver is attached to the diaphragm superiorly by the coronary ligament, which forms the right and the left triangular ligaments laterally, the falciform ligament at the anterior surface, and the gastrohepatic and hepatoduodenal ligaments at the porta hepatis. The porta hepatis is an anatomically important site as it is the region at which the portal vein, hepatic arteries and hepatic ducts enter and leave the liver respectively[39].

Conventionally, the liver is divided into the hemilivers, sectors, and segments, where the right and left hemilivers are separated at Cantlie's line (an imaginary line from the middle of the gallbladder fundus inferiorly to the superior portion of the inferior vena cava superiorly), the sectors represent the areas between the two hepatic veins or between a hepatic vein and the edge of the liver, and the segments are regions of parenchyma having independent vascular inflow, outflow, and biliary drainage[39]. There are 8 such independently acting segments also termed Couinaud segments, the knowledge of which aids segmental liver surgery in order that multiple metastases throughout distant vascular territories may be excised without compromising liver function[39].

Histologically, liver segments are divided into hexagonal lobules consisting of hepatocytes, assembled in plates radiating from a terminal branch of the hepatic vein, and the sinusoidal spaces between them[39]. At the peripheral corner of each lobule, branches

of the hepatic artery, portal vein, and bile duct are arranged to form a portal triad[38]. Blood flows through the sinusoidal space from the portal triads towards the central vein enabling mixed blood from the portal venous system and the hepatic artery to drain back to the systemic venous circulation after traversing the sinusoid. The sinusoid is lined with hepatic sinusoidal endothelial cells (HSECs) which are highly fenestrated, allowing portal blood to make contact with the hepatocytes[38]. Hepatocytes, which are arranged in cords of cells surrounding individual sinusoidal capillaries, perform the major metabolic liver functions, including protein synthesis, glycogen storage, and detoxification[40]. Other types of specialised hepatic cells include Kupffer cells, which represent resident liver macrophages[41], stellate cells (also called perisinusoidal cells or Ito cells) responsible for pathological liver remodelling (e.g. upon liver injury)[42], and bile duct epithelial cells responsible for the secretion of bile into the bile canaliculi[43].

Portal vein carries 75% of the total blood flowing into the liver; the remaining 25% is oxygenated blood transported by the hepatic artery[39]. The blood mixes in sinusoids, comes to the sinusoidal capillaries, supplies the hepatocytes, and finally drains into the central vein, which converge into the hepatic veins returning the deoxygenated blood into the inferior vena cava[39]. Hepatic blood flow rate reaches approximately 1.5 litres per min.

The liver is innervated by autonomic nerves aligned with the hepatic artery, portal vein, and bile ducts. The sympathetic innervation is postganglionic and originates in the celiac and superior mesenteric ganglia receiving the preganglionic fibres from the intermediolateral column of the spinal cord (T7-T12)[44]. The parasympathetic nerves branch off the anterior and posterior vagal trunks[44]. The anterior plexus forms a network

of nerves surrounding the hepatic artery while the posterior plexus is mainly located around the portal vein[45].

The liver serves a multitude of functions which can broadly be classified as metabolic, synthetic, detoxification and immunological. The liver also plays an important role in the regulation of haematopoiesis through the phagocytic lysis of red blood cells; a function performed by Kupffer cells to produce heme and globin[41]. Heme is further metabolised to produce iron, which is transported by transferrin to the bone marrow for haematopoiesis and biliverdin, which is ultimately secreted as bilirubin. The liver is responsible for both synthesis of several amino acids as well as their assembly into proteins; it also performs post-translational modification and metabolism or numerous protein types[46]. In addition, the liver performs the synthesis of fatty acids, triglycerides, cholesterol, and lipoproteins; it also metabolises circulating lipoproteins taken up through receptor mediated endocytosis by hepatocytes[47].The liver is also responsible for carbohydrate metabolism; both synthesising glycogen from glucose through glycogenesis and breaking down glycogen into glucose via glycogenolysis, thereby playing an important role in the maintenance of blood sugar[48]. In addition, the liver is able to synthesise glucose from amino acids, lactate, or glycerol providing an important source of glucose during times of starvation[48]. Metabolism of biochemical substances is conducted through biotransformation, which is primarily conducted by cytochromes P450 and NADPH-cytochrome c reductase, and via subsequent conjugation with glucuronide, sulphate, glutathione or acetyl adducts performed by the corresponding transferases[49].

1.3. The biological basis of colorectal cancer liver metastases

The metastatic cascade is a complicated multistep process resulting in colonisation of distant organs[50]. Microenvironment of such organs is very different from the original tumour site, and this could be both beneficial or disadvantageous for migrating cancer cells[51]. During metastatic dissemination, the overwhelming majority of travelling cancer cells perish due to multiple factors, including shear stress-related death, apoptosis or recognition followed by subsequent elimination by the host immune system[52,53,54]. In addition, some metastatic cells die during extravasation or intravasation, as these processes are associated with high degree of physical stress[55,56]. Even those cells that have successfully metastasised are prone to death at the metastasis site or indefinite quiescence/dormancy[57–59]. Thus, only a small proportion of metastasising cells are able to survive and give rise to distant colonies.

The major yet incompletely understood question is a differential tropism of metastasizing cancer cells to various organs. It has long been recognised that the distribution of metastases in cancer patients is not a random process[60,61]. For example, colon cancer most frequently metastasises to the liver[62,63], whereas lung cancer predominantly disseminates to the bone, other lung and brain[64]. Whilst there may be gross anatomical explanations for such findings (e.g. the liver receives the entire venous drainage of the gastrointestinal system via the portal vein), it has been proposed that certain microenvironmental factors, including those related to the extracellular matrix (ECM), determine the likelihood of a cancer cells ability to colonise specific organs[65,66]. For example, such factors could be the influence of resident immune cells of an organ[67] or resident fibroblasts, secreting tumour promoting molecules such as growth factors, cytokines or chemokines[68]. In the liver, for example, this could be determined by Kupffer cells[69], hepatic stellate cells[70] or resident natural killer cells[71]. Additionally, the extent of

vascularisation of an organ could also play a role in initial steps of cancer cell spreading[72,73]. A combination of these factors could define the successful seeding of cancer cells, facilitating adhesion, migration, and angiogenic switch ultimately leading to the progression of colonies. This theory, initially coined by an English surgeon Stephen Paget[74], was called the "seed and soil" theory, with metastasizing cancer cells being the "seeds", and the metastasis-favouring microenvironment being the "soil". An alternative theory, called the mechanistic theory, was proposed by an American pathologist James Ewing, and suggested that purely mechanistic factors are responsible for organotropism of different cancers. Such factors as proximity between organs, microvessel density and size, and the rate at which blood flows through them were proposed to be a predominant reason for certain organotropism of various cancers. Currently, these theories are not considered mutually exclusive, and both are likely to be relevant for determining metastatic tropism of tumour cells.

With regards to the liver, it is a highly vascularised organ, with around 30% of its volume being accounted for by blood[75]; therefore, it serves a naturally accommodating environment for travelling cancer cells to metastasise. Anatomical and physiological features of the liver determine a considerably lower pressure gradient of blood in its sinusoids in comparison with other organs (5 vs. 115 mmHg), making it an ideal target for metastasizing colon cancer cells to attach to the sinusoidal endothelium and subsequently seed[76,77]. This is largely due to the fact that the liver has a venous inflow as well as a venous outflow which no other organ has; this is what makes it a high flow, low pressure venous system.

An additional feature of the liver is the high tolerance of its intrinsic immune system[78]. The liver accumulates various antigens and pathogens coming with food in order to avoid hypersensitivity, and it's immune function evolutionary progressed towards a high tolerance phenotype. Metastasizing tumour cells could therefore exploit this feature in order to survive at the early steps of metastatic cascade. Thus, the low blood flow and relatively weak local immunity define the liver as an ideal "soil" for metastasis. In addition, the colon and liver are localised in tight proximity with each other, having partially common vascular beds, which could also explain colon cancer organotropism from the mechanistic point of view.

Having detached from the primary location at the bowel wall, colon cancer cells intravasate and travel towards the hepatic sinusoid. Although the majority of these cells die on their way to the liver, combined microenvironmental and mechanistic forces described above facilitate the survival of a small proportion of these cells which will eventually give rise to metastases. There are several hypotheses of how travelling cancer cells physically secure their place in the liver. Some suggest that cancer cells get trapped in the hepatic sinusoids and therefore become able to adhere[79]. Others speculate that cancer cells can actively adhere to HSECs through ligation with intercellular adhesion molecule 1- (ICAM) and vascular cell adhesion molecule 1- (VCAM) mediated manner[80]. In either event, the adhered cells may leave the vascular space by means of extravasation, or alternatively, they may proliferate within the sinusoidal space rather than invading[81]. Migration of cancer cells into liver parenchyma occurs through transcellular or parallel diapedesis, whereby tumour cells induce apoptosis of endothelial cells or travel between adjacent endothelial cells, respectively[82]. As a result of this process, liver parenchyma becomes colonised by metastasizing cancer cells in four

consecutive steps: microvascular, intralobular preangiogenic, panlobular angiogenic and lobar growth phases[83]. Briefly, these steps are characterised by settlement and proliferation of cancer cells in the intrasinusoidal space, the space of Disse (perisinusoidal) or periportal locations. Subsequent outgrowth of these lesions is typically triggered by the angiogenic switch, a process enforced by infiltrating myeloid-derived suppressor cells producing pro-angiogenic factors such as vascular endothelial growth factor (VEGF) or fibroblast growth factor 2 (FGF2) in the tumour microenvironment[63]. Developing liver metastases exhibit several characteristic growth patterns[84], such as the replacement pattern, which leaves the architecture of the liver intact thereby making such lesions difficult to detect; desmoplastic pattern, characterised by extensive stromal infiltration, fibrotic response and considerable ECM deposition; and pushing pattern, in which the hepatic architecture becomes completely lost upon tumour outgrowth[83]. There are multiple factors that may determine the pattern in which metastases will develop, including the initial space of colonisation, microenvironmental factors mediated by fibroblasts, Kupffer cells and hepatic stellate cells as well as the hypoxic status of growing tumour colonies.

To conclude, our understanding of the biological basis of colorectal cancer liver metastases is still fragmental; however, there are clearly several hallmarks of this process which require in-depth investigation. One of these hallmarks is the ECM, which could significantly affect the fate of metastasizing cancer cells, being either a barrier or favouring factor. Thus, we turn our attention do the role of the ECM in the development of hepatic metastases.

1.4. Extracellular matrix in cancer biology

Recently, a large-scale genetic study established four molecular subtypes of human colorectal cancer[85]. The immune subtype is characterised by microsatellite instability and strong immune activation, the canonical subtype demonstrates an epithelial phenotype and activation of Myc and Wnt pathways, whereas the metabolic subtype has a marked metabolic deregulation. Finally, the mesenchymal phenotype, affecting 23% of patients, is characterised by a considerable ECM deposition, activation of transforming growth factor β (TGF-β) pathway and worst disease-free survival as compared to all other subtypes. This study indicates that ECM component is important in at least a quarter of patients with colorectal cancer and highlights the significance of this factor in a subset of patients with the worst survival rate, specifically those who develop liver metastases. Importantly, the above mentioned bioinformatics research was also supported by extensive basic studies performed in mice[86].

The ECM constitutes a major part of all multicellular organisms[87]. It represents a complex meshwork consisting of numerous classes of proteins with pleiotropic functions. Being a non-cellular component, the major role of ECM is to maintain structural and mechanical support for a tissue[88]. Furthermore, it controls the integrity and morphology of a tissue through regulating behaviour of cells through a complicated network of cell-matrix interactions[89].

Accumulating evidence implicates the ECM as an important player in tumour development, including cancer progression and metastatic dissemination[90,91,92]. Briefly, changes in tumour ECM augment cell motility and invasiveness, resulting in the loss of cell adhesion[93]. Subsequently, this is followed by the entry of cancer cells into the circulation successfully paving their way through the dense matrix. Eventually, this leads

to dissemination of these cells into distant organs to form metastases[94]. Additionally, ECM may play a role in the onset of certain cancer types; for example, pathological fibrosis in the liver and breast can increase the risk of hepatocellular carcinoma and breast cancer, respectively[95,96].

The ECM is predominantly formed from three main classes of proteins, namely collagens, proteoglycans and glycoproteins[97,98]. These are fibrillar and cross-linked molecules, contributing to the insolubility, rigidity and porosity of ECM, factors that can be altered by the activity of another class of molecules called ECM regulators. The ECM can rapidly change its stiffness from rigid to loosen and vice versa due to ECM regulators[99]; these consist of a variety of extracellular proteases, their inhibitors, growth factors, cytokines, mucins, galectins etc. and participate in crosslinking, degradation, re-depositing, stiffening and remodelling of ECM constituents. The subtle balance between these proteins determines the state of ECM, which can be changed by cellular signals if necessary. Importantly, ECM remodelling by these regulators is recognised as one of the crucial steps in metastatic process as well as in response to therapy, as demonstrated by studies implicating ECM stiffness in promoting metastatic cancer cell interaction with the endothelium[100] and altering the effectiveness of chemotherapy[101,102]. Furthermore, interaction with key ECM protein fibronectin promoted resistance to radiotherapy in lung cancer cells[103]. Interestingly, certain ECM components such as high-molecular-mass hyaluronan, were reported to promote cancer resistance[104].

Collagen is the most abundant ECM protein in the body, and its excessive deposition has been shown to be important for cancer development. Accumulating concentrations of collagen in the tumour microenvironment result in increased

mechanical force within the tissue[105,106]. This leads to changing permeability of the basal membrane through altering its viscosity and enabling various macromolecules and cells to migrate in and out[107]. Additionally, overexpression of collagen in tumour microenvironment may lead to deregulated cell polarity, altered cell-cell adhesion, and growth factor signalling, cumulatively facilitating cancer progression and metastasis[108,109]. Moreover, it was reported that collagen cross-linking mediated by lysyl oxidase in tumours promoted matrix density and hypoxia through a reduction in angiogenesis[90,110]. Whilst cancer patients overexpressing lysyl oxidase have generally poorer survival[111], inhibition of this protein inhibited fibrosis and increased tumour latency[112]. Another study indicated that the collagen-mediated ECM deposition altered the immune microenvironment towards a pro-M2 phenotype which promoted tumour progression[113]. In addition, collagen signals to cells through the receptor Tyr kinases Discoidin domain-containing receptors (DDR) 1 and 2[114], both involved in cancer development through promoting migration[115] and epithelial-to-mesenchymal transition[116] required for metastasis. On the other hand, degradation of tumour ECM by proteases has also been reported to contribute to cancer cell invasion[105]. In particular, collagen cleavage by matrix metalloproteases (MMPs) was linked to the formation of a looser ECM, facilitating migration of cancer cells[117,118]. Matrix cleavage by MMPs and ADAMs is well reported to liberate pro-angiogenic growth factors bound to the ECM[119], bringing another fundamental role of the ECM and demonstrating how regulation of its deposition and/or breakdown can promote cancer progression. Thus, both increased stiffness and excessive degradation of ECM components, primarily collagen, can stimulate metastatic dissemination of cells and therefore the disruption of a fine balance of ECM regulator enzymes may lead to tumour invasiveness.

1.5. Experimental models of colon cancer and liver metastases

Currently, the experimental research of primary or metastatic colon cancer is based on the use of animal models of this disease. Whilst *in vitro* cell culture experiments have been invaluable for cancer research, this approach is limited in several critical ways[120]. Results generated *in vitro* can rarely be recapitulated in human disease due to several reasons, including ignorance of the impact of tumour microenvironment as well as the complexity of crosstalks between the immune and other systems of the body and tumour tissue. It is important to describe the main animal models of colon cancer and associated liver metastases in order to have a better understanding of this area of research.

1.5.1. Graft models of colon cancer

Several graft models of colon cancer exist. The classical method of subcutaneous injection of colon cancer cells is commonly used; however, it may not be physiologically relevant as it ignores the microenvironmental component of the colon, which has a strong impact on disease progression[121]. It could be successfully used, however, to assess the viability of cells upon the *in vivo* administration of various drugs or radiotherapy[122]. The major advantage of the subcutaneous injection is its relatively low cost and technical ease of this model. It does not require complicated equipment or tools and can be performed in the majority of research laboratories. In addition, subcutaneous injections are characterised by a rapid tumour growth: two weeks is generally enough to grow a 300-500 mm^3 xenograft (depends significantly on cell type used). Subcutaneously injected colon cancer cells rarely metastasise to distant organs.

Alternatively, colon cancer cells can be injected directly into the caecal wall[123]. This method results in the development of multiple bowel tumours at the site of injection. Although this technique is technically challenging, it provides a reliable model to study primary colon cancer and with use of the appropriate cell line can recapitulate all stages of the metastatic process. An alternative to this method is a transplantation of a mass of subcutaneous tumour into the caecum wall. The resected subcutaneous tumour grown in another mouse should be cut into multiple small pieces and carefully injected into the bowel wall to ensure the growth of xenograft[123].

To generate liver metastases from colon cancer, the well-recognised technique of an intrasplenic injection is commonly used. In this model, tumour cells are injected into the mouse spleen in order to simulate the metastatic dissemination through the portal vein into the liver parenchyma. This technique develops multiple liver metastases in 2-4 weeks and works well for the majority of colon cancer cell lines. It is a relatively straightforward method, albeit requiring some practice to perform the injection carefully to avoid the leakage of tumour cells, which would otherwise result in development of unwanted peritoneal tumours. An alternative liver metastasis method is the injection of colon cancer cells directly into the liver parenchyma. Such intrahepatic orthotopic injection is highly challenging technically due to the fact that the liver is extremely vascularised, and any minor damage could cause a significant blood leakage. Despite this model ignores the step of metastatic seeding into the liver, it develops a single well-defined lesion, growth of which could easily be monitored by imaging systems, such as computed tomography (CT) or magnetic resonance imaging (MRI), which is nearly impossible to do in the intrasplenic injection model. In addition, intrahepatic injection enables to develop

tumours in 1-2 weeks, which is rather advantageous over the intrasplenic injection technique.

It is important to note the developing of tumours using all the above-mentioned models can vary depending on the cell line and mouse species. Some of the cell lines, like MC38, are readily forming tumours in any of the above-mentioned models, however, there are cell lines which are very difficult to grow as *in vivo* xenografts (e.g. DLD1).

1.5.2. Chemically induced models of colon carcinogenesis

The main advantage of using chemically induced models of colon carcinogenesis is their great reproducibility; however, such models generally require a long-term exposure to genotoxic agents and thus experiments might last for months. A wide range of carcinogens are used for the development of colon cancer in mice and rats. The earliest model of this kind was established more than 70 years ago and involved oral administration of polyaromatic hydrocarbons, specifically 1,2,5,6-Dibenzanthracene and 20-Methylcholanthrene to mice, which subsequently developed intestinal carcinoma[124]. Several years later, Walpole et al. showed that continuous (5 days a week) subcutaneous injection of 4-Aminodiphenyl (3.6-5.8 g/kg) or 3:2'-Dimethyl-4-Aminodiphenyl (2.8-35 g/kg) resulted in the development of spontaneous small and large intestinal tumours along with lesions in the salivary or sebaceous glands in 259 to 500 days[125]. Similar effects of were further demonstrated by Reddy and Ohmori[126], who showed that weekly 50 mg/kg subcutaneous injections of 3:2'-Dimethyl-4-Aminodiphenyl led to the development of colon, small intestinal, ear duct, and skin tumours within 20 weeks of treatment.

Interestingly, animals fed with a high-fat food had a much higher incidence of tumours compared to those having a low-fat diet (74% vs. 30%, respectively).

Besides polyaromatic hydrocarbons, multiple studies indicated the role of heterocyclic amines, in particular, 2-amino-1-methyl-6-phenylimidazo[4,5-b]-pyridine (PhIP), in colon carcinogenesis. Heterocyclic amines are generated during meat or fish broiling as by-products of heating of creatine, monosaccharides, and some amino acids[127]. Long-term exposure to heterocyclic amines primarily cause cancers of the colon, mammary gland and prostate[128]. In the study by Hasegawa[129], the addition of PhIP at a dose of 100 part per million (ppm) in the diet for up to 104 weeks resulted in 43% of rats developing colon adenocarcinomas. However, none of the rats receiving 25 ppm of PhIP developed colon tumours in this study. In another study[130], the addition of 400 ppm of PhIP in the diet of F344 male rats resulted in 50% of animals developing tumours at 52 weeks. Notably, administering the high-fat diet in combination with PhIP enabled much more frequent onset of tumours[131]. PhIP was reported to induce adenomatous polyposis coli (*Apc*)-mutated tumours with microsatellite instability, although resulting tumours had no mutations in *K-ras* or *p53*[132]. Compared to other carcinogens, however, PhIP is not considered as a good option due to relatively high price of this compound. Another common carcinogen belonging to heterocyclic amines, 2-amino-3-methylimidazo[4,5-f]quinoline (IQ), was shown to induce intestinal and liver tumours in rats after 58 weeks upon the addition to food at a dose of 125 ppm[133]. However, it was also reported to induce forestomach and lung tumours[134], indicating its lower specificity for the development colon cancer.

Most of more recently developed models of chemically induced colon cancer involved the use of dimethylydrazine (DMH) or its metabolites, primarily due to its higher potency as compared to other compounds. The pioneering research in this direction was performed by Laqueur et al.[135], who first discovered that consumption of cycad nut food obtained from the plant *Cycas circinalis L.* indigenous to Guam, leads to development of intestinal, kidney and liver tumours in rats. Further research demonstrated that cycad meals contain the carcinogen cycasin, which in metabolic conditions hydrolyses to a highly reactive genotoxic agent methylazoxymethanol glycoside (MAM)[136]. The metabolite of DMH, MAM, yields a methyl diazonium ion which can alkylate nucleic acids, in particular, guanine[137,138], therefore inflicting abconsiderable DNA damage. Importantly, DMH demonstrates superior results as compared to other carcinogens used for colon cancer development. In the study by Femia et al.[139], almost 100% incidence of colon tumours was observed in male F344 rats after just 28 days of 2 subcutaneous injections of DMH (150 mg/kg) one week apart.

Another well-recognised carcinogen, azoxymethane (AOM), also metabolises to MAM, thus being able to reproducibly induce colon cancer in mice[137]. Tumours resulting from DMH and AOM treatment exhibit mutations in *K-ras* (66% and 33% of animals, respectively)[140,141] and *β-catenin*[139,142]. In addition, AOM was reported to induce colon cancer with the mutation in the *Apc*[143] and *TGF-β* Receptor II[144] genes. Importantly, both DMH- and AOM- induced tumours exhibited histopathological resemblance to human colon cancer, making these models especially useful for experimental cancer research[132].

To summarise, there are multiple chemicals available to reproducibly induce colon cancer in mice and rats; however, the generation of such tumours might require a considerable amount of time and effort.

1.5.3. Genetically engineered models of colon carcinogenesis

In colon cancer, the Wnt/β-catenin signalling plays a major role in maintaining the intestinal epithelial homeostasis [145]. Downstream targets of Wnt, β-catenin and Apc, are frequently deregulated in patients with colon cancers and FAP. Apc acts as a tumour-suppressor, as *Apc*^{mut} mice developed in 1990 exhibited the occurrence of multiple adenomas throughout the entire intestinal tract at an early age[146]. In the next decades, 11 genetically engineered mice harbouring the *Apc* gene mutation were generated, all characterised by hyperproliferation of intestinal glands resulting in the development of polyps in the duodenum, small intestine, large intestine and rectum[147]. However, despite being a reproducible source of tumours, all *Apc* mouse models mostly develop lesions in the small intestine, which is in contrast to human disease, which predominantly occurs in the colon. Metastases to distant organs are rarely reported for *Apc* mouse models.

Another model affecting the Wnt signalling resulting in colonic tumour formation is the mutant β-catenin mouse model. Romagnolo et al.[148] generated transgenic mice which exhibited an increase in the expression of a NH2-terminally truncated mutant β-catenin in the intestine. After 28 days of age, such mice developed multifocal dysplastic lesions in the small intestine, similar to those observed in the FAP mouse models. Some of the mice also developed dysplasia of low grade in the duodenum and ileum. The observed intestinal lesions where characterised by abnormal morphology with increased

apoptosis and proliferation. Similar to *Apc* mouse models, the colon was not affected in these mice[148]. Interestingly, activation of β-catenin in these mice resulted in the development of severe polycystic kidney disease. Another β-catenin activated mouse model was developed by Harada et al. [149] In this study, intestinal polyps were determined at highest density in the duodenum, proximal jejunum, and ileum. No lesions were found in the colon. Histopathological examination revealed that the resulting tumours were very similar to those generated using the *Apc* mouse models.

Apart from the Wnt/β-catenin pathway, an important role in colon carcinogenesis belongs to Ras mutations. Constitutive activation of *K-ras* oncogene is observed in around 50% of patients with colorectal adenocarcinoma[150]. In one genetic mouse model, overexpression of *K-ras* gene in mice caused early dysplastic events in colonic crypts (so-called aberrant crypt foci), as well as lung adenocarcinoma, skin papilloma and thymic lymphoma[151]. Another group generated transgenic mice with overexpression of *k-ras* gene specifically in the intestine[152]. More than 80% of such mice exhibited single or multiple lesions at 2-20 months of age, including aberrant crypt foci and adenocarcinomas of the periampullar region (3%), duodenum (46%), jejunum (44%), and colon (7%). Interestingly, some of these mice also developed a spontaneous mutation in the *p53* gene, suggesting a crosstalk between K-ras and p53.

TGF-β is a pleiotropic cytokine regulating cell differentiation, apoptosis, and proliferation. In cancer, TGF-β is generally considered as a tumour suppressor, and around 25% of colorectal adenocarcinomas exhibit a mutation in the *TGF-β* gene or members of its pathway[153]. Because TGF-β mutations in mice are generally lethal, several attempts to target factors downstream TGF-β have been performed. For example, Smad

family of proteins are phosphorylated by the TGF-β and translocate to the nucleus in order to activate transcription of multiple genes. Genetically engineered mice with a knockdown of *Smad3* gene have been shown to develop aggressive colorectal adenocarcinomas as well as multiple metastases in the lymph nodes[154]. Interestingly, when these mice were housed in specific pathogen free conditions, none of them developed colon tumours, suggesting the importance of gut microbiota in the TGF-β driven carcinogenesis[155].

Overall, there are multiple methods currently available to generate primary colon tumours, but not liver metastases in mice. All of the above-mentioned methods have their own pros and cons and depending on the experimental setup these methods can be employed with different effectiveness. Therefore, careful optimisation including pilot experiments must be performed in order to achieve better results. It is also important to note that the results generated using these methods should be carefully analysed to exclude potential sources of bias. For example, there are several approaches to assess the tumour burden, including weighing of the resected cancer tissue, immunohistochemistry analysis of tumour area, visual counting of macroscopic colon cancer lesions or liver metastases, and disaggregation of tumour tissue followed by fluorescence activated cell sorting (FACS) analysis of fluorophore positive cancer cells (for this purpose, genetically engineered cells expressing a fluorescent protein should be used). Frequently, these methods of tumour burden assessment can yield variable results as they are not directly comparable; the number of macrometastases, for example, can be different from the number of micrometastases due to peculiarities of the metastatic growth of a given cell line. Further, the amount of water in the tumour tissue can affect its weight, therefore,

all resected tumours should be dried on a napkin for a certain amount of time before being placed on the scales.

To conclude, mouse models of colon cancer are practiced extensively throughout the world today; these models have provided tools to study in depth the molecular biology and biochemistry of colon cancer and associated liver metastases, and their use greatly facilitates cancer research.

1.6. Post-translational modifications of tumour ECM

ECM in cancers is notable for its disorganisation, with altered expression of its components in comparison with normal tissue counterparts. Moreover, multiple post-translational modifications can significantly affect the interactions between ECM proteins and cancer or stromal cells. For example, lysyl oxidase-mediated crosslinking of collagen has been reported to stiffen the ECM, promoting fibrosis, activating focal adhesions, and enhancing progression of breast tumours[112,156]. Further, proteolytic cleavage of various ECM components via MMPs is well documented to remodel the tumour matrix and alter tissue architecture, generally leading to changes towards more aggressive tumour phenotype[157,158]. Glycosylation of certain ECM molecules, such as fibronectin, was reported to induce epithelial-to-mesenchymal transition (EMT) in prostate cancer cells[159] and also was associated with invasiveness in bladder cancer[160].

Accumulating evidence implicates the post-translational modification of ECM proteins as crucial for cancer development[161]. However, physiologically significant post-

translational modifications including deamidation, carbamylation, or citrullination, have been infrequently described in cancer biology.

1.6.1. General description of citrullination

In 1914, the free amino acid citrulline was extracted from the watermelon by Koga and Ohtake[162]. Citrulline is a non-coding amino acid and can only be produced in the body through a posttranslational modification of proteins via conversion of arginine residues. The reaction of citrullination is catalysed by peptidyl arginine deiminase (PAD) enzymes, which hydrolyse the arginine guanidinium into a urea group, resulting in 1 Da change in molecular mass. Citrullination occurs in the cell nucleus and cytoplasm, yet extracellular space proteins are also known to be citrullinated. The list of the most commonly citrullinated proteins found almost in every study includes Vimentin, Fibronectin, Keratin, Tubulin, Actin, Filaggrin and various Histones. The biological role of citrullination is incompletely understood, although there is evidence that citrullination affects transcription through modification of transcription factors and histones[163,164]. Furthermore, whilst arginine is a positively charged amino acid, citrulline has a neutral charge. Because of this, citrullination can increase protein hydrophobicity, resulting in substantial changes in protein folding, and, as a consequence, altered function[165]. This is especially seen in arginine-rich proteins, where the effect of citrullination is even more dramatic. Finally, citrullination is also important for neutrophil extracellular trap (NET) formation and generation of neoantigens, each of which will be described in detail below.

Interestingly, the free citrulline can also be found in the body as a by-product of the urea cycle[166]. In mitochondria, citrulline production is catalysed by ornithine transcabamoylase from ornithine and carbamoyl phosphate. The resultant citrulline and

aspartate are condensed to form argininosuccinate, catalysed by argininosuccinate synthetase. The plasma concentration of citrulline can therefore be used to diagnose inborn errors of ammonia detoxification or arginine synthesis.

1.6.2. Protein arginine deiminases as drivers of citrullination

The role of PADs in catalysing citrullination was first described in 1977[167] yet the catalytic mechanism of this reaction was only recently described[168]. PADs represent a family of five (PAD1, PAD2, PAD3, PAD4, and PAD6) calcium-dependent isoenzymes grouped as a 355-base pair cluster on chromosome 1 (1p36.13). Phylogeny studies have revealed that all PAD isotypes developed as a series of gene duplications in a common ancestor[165]. All human PADs share a considerably high degree of homology between each other and also with orthologous proteins of mammalian vertebrates (70-95% identical amino-acids). Their distribution in tissues, however, is highly variable (Table 1), with PAD2 being the most widely expressed isoenzyme.

Table 1. Expression of PAD genes in the human body

Gene	Expression in the human body[169,165]
PAD1	All living layers of the epidermis, hair follicle, uterus
PAD2	Skeletal muscle, brain, immune cells, skin, peripheral nerves, uterus, spleen,

	secretory glands, pancreas, kidney, inner ear
PAD3	Hair follicle, skin, peripheral nerves
PAD4	Immune cells, brain, uterus
PAD6	Ovary, egg cells, embryo

As demonstrated by functional and structural studies, all PADs are able to bind Ca^{2+} ions in order to be activated. Three molecules of Ca^{2+} can bind to one molecule of PAD at specific EF-hand domains of the protein[165]. Interestingly, there is a prokaryotic enzyme able to citrullinate proteins is expressed in the gram-negative pathogenic bacterium *Porphyromonas gingivalis.* This enzyme, named AAF06719, is evolutionary unrelated to PADs and its activity is not dependent on calcium ions[165].

1.6.3. Roles of citrullination in inflammatory conditions.

Citrullination is best studied as a pathological entity in the setting of inflammatory disease. It is thus well recognised that citrullination is an inflammation-associated phenomenon which occurs in multiple tissues and organs[170]. Under certain conditions citrullinated proteins can inflict the production of autoantibodies against these proteins. This has been documented in several autoimmune or inflammatory pathologies including periodontitis[171], autoimmune encephalomyelitis[172], and rheumatoid arthritis (RA)[173]. The resulting production of anti-citrullinated protein antibodies (ACPAs) leads to a local information response and exacerbates the severity of disease. It is likely that citrullinated proteins *per se* do not act as the initial immune trigger; rather, they uphold and intensify

an established inflammatory cycle activated by environmental factors, such as acute bacterial infection or toxin exposure or other immunogenic compounds. Accumulating research evidence indicates that protein citrullination caused by an initial inflammatory trigger leads to the generation of neoantigens that significantly enhance the T-cell mediated immune response. Smoking for example, demonstrates a dose-dependent association with the presence of anti-citrulline antibodies in RA patients[174], whilst following immunisation using protein immunogens, $CD4^+$ T cells become able to recognise citrullinated proteins[175]. With regards to specifically the ECM, several studies demonstrate that in RA patients, citrullinated aggrecan and vimentin epitopes bind human leukocyte antigen (HLA) molecules, promoting a proinflammatory response[176,177,178]. These data directly implicate citrullination of ECM proteins in the inflammatory pathophysiology of RA. Finally, the concentration of autoantibodies against a citrullinated B cell epitope on α-enolase were markedly increased in RA patients[179]. Taken together, these data demonstrate that citrullination of various proteins results in the formation of neo-antigens and is directly involved in the development of an inflammatory response.

Although citrullinated proteins are detectable in almost any inflamed tissue or organ[180], the characteristic autoimmune response mostly occurs in the synovial tissue of patients with RA. The formation of ACPAs is very specific for this disease and also can precede clinical manifestations of RA. Up to 80% of individuals with RA have positive serum titers for ACPAs. Due to high specificity and sensitivity (98% and 80%, respectively), they are currently used as biomarkers of this disease in the clinic. Other markers of RA, such as, for example, rheumatoid factor (RF), are much less sensitive for the detection of this disease (approximately 70%)[181]. Clinically, the most commonly

utilised commercial assay currently involves a synthetic cyclic citrullinated peptide (CCP2) and has a superior sensitivity for the detection of ACPA[173]. However, ACPAs are also known to be present in some other inflammatory conditions (e.g. autoimmune encephalomyelitis, periodontitis), and moreover, can be detected in 2% of healthy individuals[182]. It should also be noted that in a very small cohort of RA patients citrullinated proteins do not elicit an autoimmune response, probably due to hormonal, genetic, and other factors[180,183].

Besides its role as a biomarker, little is known about whether citrullinated proteins are targeted by ACPAs in affected joints and if they have any impact on the pathophysiology of this disease. ACPAs are usually found in the blood of individuals who will develop RA in the future, implicating citrullination in the early steps of disease development and progression. High levels of ACPAs also correlate with more aggressive disease course[184], extra-articular manifestations[185], coronary artery calcification[186], and more severe joint destruction[187] in RA patients. In support of clinical findings, animal models demonstrated that the administration of citrullinated proteins exacerbates collagen-induced arthritis[188]. Similarly, treatment with pan-PAD inhibitors Cl-amidine or BB-Cl-amidine abrogated collagen-induced arthritis in mice and exhibited altered T-cell mediated immune responses[189,190]. Several studies in PAD4-deficient mice demonstrated a marked amelioration of collagen-induced arthritis severity and reduced levels of myeloid cells, T_H17 cells and serum interleukin 6 as compared to control animals[191,192]. Collectively, these data indicate that citrullination-mediated formation of ACPAs influences RA development and therefore citrullination can be considered not only as biomarker but also as a therapeutic target for this disease.

Chromatin citrullination has recently been implicated in the formation of NETs via a specific cell death mechanism termed NETosis. Here, dying neutrophils release decondensed DNA into the ECM, disarming bacterial virulence factors and killing bacteria through neutrophil-derived anti-microbial proteins that have affinity to DNA[193]. In patients with RA, isolated neutrophils demonstrated spontaneous NET formation *in vitro*[194,195]. Cell-free DNA characteristic of NETosis was present at higher concentrations in the serum from RA patients compared with healthy subjects[196]. Further, IgG fractions derived from blood of RA patients with high titers of ACPA and a well-recognised RA autoantibody RF markedly promoted NETosis in control or RA neutrophils without additional priming by cytokines[197]. Another study *in vivo* showed that NETs promoted the development of T_H1 immune response in collagen-induced arthritis through maturation of dendritic cells[198]. In addition, NETs are especially relevant in cancer; in particular, they were recently shown to be required for colorectal cancer liver metastases seeding[199,200]. Thus, these data indicate that citrullination-dependent event NETosis is important for the pathogenesis of RA and cancer.

Besides RA, citrullination has also been implicated in several other inflammatory disorders. Patients with systemic lupus erythematosus (SLE) were reported to have a decreased ability to degrade NETs[201], which were also suggested to be a major source of autoantigens in this disease[202,203]. Furthermore, SLE patients exhibited enhanced NET formation in isolated neutrophils as compared to healthy control neutrophils[204]. In lupus-prone mice, inhibition of PADs by Cl-amidine disrupted NET formation and protected against kidney and vascular inflammation[205]. Citrullination and NET formation also contribute to the pathogenesis of atherosclerosis[206], thrombosis[207], and inflammatory bowel disease[208].

1.6.4. Roles of citrullination in neurological disease

Demyelination of central neurones is responsible for the clinicopathological features of multiple sclerosis (MS). Myelination is mediated by proteins such as glial fibrillary acidic protein (GFAP) and myelin basic protein (MBP), the latter of which is required for successful layering of the myelin sheath, insulation of the neurone and thereby successful signal conduction at the nodes of Ranvier. Interestingly, both GFAP and MBP are extensively citrullinated in patients with MS[209,210] and citrullination is proposed to target these proteins for protease-mediated degradation, abrogating myelination and generating neoantigens[211]. MBP citrullination also weakens its interactions with phospholipids, reducing adherence between layers of the myelin sheath and ultimately resulting in demyelination[212]. Notably the functional significance of GFAP citrullination is unknown, although citrullinated GFAP promotes the disassembly of the intermediate filament, a structure important for guiding radial axon growth[213]. Consistently with high levels of citrullination, the brains of patients with MS exhibited higher PAD concentration and activity[214]. Interestingly, there was a reduction in methylation of the PAD2 promoter in MS tissues compared with those from normal brain, implicating epigenetic regulation in promotion of citrullination in MS patients[215]. In keeping with these findings, mice overexpressing PAD2 under control of the MBP promoter exhibited high levels of demyelination along with clinical evidence of MS[216].

Citrullination has also been linked to other neurodegenerative pathologies. High PAD2 and citrullination levels were detected in the brains of mice affected by prion disease[217], as well as those of patients with sporadic Creutzfeldt-Jakob disease[218]. An

increase in protein citrullination and activation of PAD2 has also been observed in the hippocampi of patients with Alzheimer's disease[219] whilst in patients with Parkinson's disease, high concentrations of citrullinated proteins were detected in the dopaminergic neurons of the substantia nigra; the neurones that undergo degeneration as part of this pathology. Finally, in an experimental model of spinal cord injury, PAD inhibition using Cl-amidine led to reduced apoptosis of neural stem cells and faster regeneration[220]. Taken together, these data indicate that PAD-mediated citrullination plays a role in the pathogenesis of neurodegenerative disorders, although in several cases, the mechanisms through which citrullination mediates disease progression remains unclear.

1.6.5. Citrullination and cancer

Given the important role of citrullination in the pathogenesis of the aforementioned inflammatory diseases and the central importance of inflammation for the development and progression of cancer, interest has emerged in the investigation of links between citrullination and cancer. Accumulating evidence implicates PAD4-driven citrullination in the development and progression of various cancers. Several studies have demonstrated that PAD4-mediated citrullination of methylated arginine residues on histones H3 and H4 reduced the expression of p53 target genes including *p21*, *CIP1*, and *OKL38*[221,222]. PAD4 also interacts directly with p53 on the *p21* gene promoter[223]. PAD4-mediated citrullination of inhibitor of growth protein 4 (ING4) also influences expression of p53 dependent genes[224], and DNA damage *in vitro* resulted in PAD4-mediated citrullination of histones H3 and H4[225]. In another study, neutrophil-derived NET triggered HMGB1 release and activated TLR9-dependent pathways in cancer cells to promote their adhesion, proliferation, migration, and invasion[200]. Further, in a mouse

model of infection using caecal ligation and puncture, microvascular NET deposition was followed by consequent trapping of circulating lung carcinoma cells within DNA webs, contributing to metastases[199]. Also, such NET trapping was associated with increased formation of hepatic micrometastases at 48 h and gross metastatic disease burden at 2 weeks following tumour cell injection[199]. In the study by Cook et al.[226] immunisation of mice with two citrullinated α-enolase peptides induced strong T_H1 responses resulting in improved survival of melanoma and pancreatic cancer xenografts. PAD2 was shown to be underexpressed in colorectal carcinoma tissues, and its genetically induced overexpression inhibited proliferation of colon cancer cells through protein citrullination[227]. Collectively, these data implicate the involvement of PADs/citrullination in carcinogenesis.

Furthermore, elevated PAD4 expression has been identified in various tumour types relative to their respective normal tissue counterparts. In a large study by Chang et al.[228], an extensive immunohistochemistry (IHC) profiling of 1673 tumours of different origin revealed the expression of PAD4 in carcinomas of the breast, lung, liver, colon, oesophagus, kidney, ovary, uterus, bladder and soft tissue. Around 40% of cells in malignant lymphomas also expressed PAD4 indicating that PAD4 expression is associated with cancer development from all embryological lineages. By contrast, various benign tumours as well as non-tumour inflamed tissues did not express PAD4[228] implicating citrullination in the progression from benign neoplasm to invasive malignancy. Several studies also demonstrate the ability of PAD4 to citrullinate transcription factors Elk-1 and p300, inducing gene expression changes in cancer cell lines[229,230]. Elk-1 has been reported to interact with BRCA1a and BRCA1b[231], suggestive of the role of PAD4 in the development of breast cancer.

With regards to other PADs, a downregulation of intratumoural PAD2 expression was identified in a cohort of 98 colorectal cancer patient specimens in comparison with normal mucosa of 50 matched healthy control tissues[232]. Similarly, IHC analysis of PAD2 cytoplasmic or nuclear staining was not significantly associated with age, survival time, oestrogen or progesterone receptor status in patients with ductal breast carcinoma in situ[233]. Thus, although early investigation implicates PAD4 in the development and progression of cancer, preliminary investigation for other members of the PAD family has as yet failed to yield positive results.

1.6.6. Means of detecting citrullination

Accurate and robust detection of citrullination is required for citrullination research; however, this is challenging for a number of reasons. First, arginine citrullination results in only a small mass change (0.9848 Da), making it difficult to detect using mass spectrometry-based methodologies. Second, citrullination is a less frequent and less readily enriched modification when compared to other post-translational modifications such as acetylation or phosphorylation. Finally, as citrullination is a reversible process, its presence may be transient and present only at specific points throughout cancer progression. Despite these difficulties, over the past two decades, various strategies have been developed to precisely quantify citrulline activity in biological samples.

The first approach, proposed by Senshu et al.[234,235,236] involved development of an antibody that specifically recognises citrullinated arginine residues modified with a

uriedo group adduct because of a reaction between citrulline and antipyrine 2,3-butanedione monooxime in the presence of $FeCl_3$ and strong acid solution. This allows detection of citrulline using standard antibody-based methods, such as enzyme-linked immunosorbent assay (ELISA) or immunoblotting using commercially available antibodies raised against 2,3-butanedione monooxime-modified citrullinated antigens.

Subsequently an antibody (F95) has been developed against a synthetic peptide antigen consisting of 10 peptidyl-citrullinates attached to two glycines and a cysteine, named deca-citrullinated protein[237]. The antibody demonstrated strong immunoreactivity to various citrullinated proteins, including histones, myelin basic protein, bovine serum albumin, and ovalbumin[237].

More recently, a commercially available chemical probe capable of detecting protein citrullination has been developed. Bicker et al.[238] synthesised a meta azide substituted phenylglyoxal tagged with fluorescent dye rhodamine that can covalently bind to citrullinated residues in the presence of 20% trichloroacetic acid. This fluorescent probe was shown to detect femtomolar concentrations of citrulline in various samples and enabled accurate quantification of gel electrophoresis-resolved bands using a fluorescent scanner[238]. Another advantage of this approach is its speed of use when compared with 2,3-butanedione monooxime modification, as labelling and analysis generally take less than 3 h.

Finally, protein citrullination can be detected and accurately quantified by various mass spectrometry workflows[239]. Modern liquid chromatography–mass spectrometry (LC-MS/MS) instruments are capable of distinguishing 1 Da mass shifts on arginine

residues, allowing the resolution of minute differences in ion mass/charge signals[240,241,242]. This enables precise quantification of every citrullination site in each protein in the sample. The major disadvantage of this method is that deamidation of asparagine or glutamine, an alternative post-translational modification, also leads to a mass shift of 1 Da thereby potentially confounding the analysis of citrullination status[243]. To account for this, manual confirmation of peptide identity and each citrullination event by checking tandem mass spectra is required for determination of citrullination status using mass-spectrometry. Further, identification of citrullination by proteomics is extremely instrument-demanding and requires the use of devices with high mass accuracy and mass resolution; therefore, only a small proportion of MS instruments are capable for citrullination analysis.

1.6.7. Inhibitors of citrullination

The recognised importance of citrullination for a variety of disease processes has led to the development of methods for the inhibition of citrullination. Inhibition of PAD enzymes is an obvious target in this regard. Because the catalytic mechanism of PADs strongly resembles that of cysteine proteases and given that irreversible inhibitors for these enzymes were successfully developed, the first highly potent PAD inhibitors were similar to cysteine protease inhibitors[244]. The highest degree of PAD inhibition is achieved by Cl-amidine, which irreversibly binds to the PADs enzymatic sites. Cl-amidine has been utilised in multiple *in vitro* and *in vivo* experiments that have helped in the understanding of PADs functional activity[208,245,246]. Cl-amidine has multiple analogues, including among others, F-amidine, H-amidine, and YW3-56[244]. The majority of these analogues are significantly less potent than Cl-amidine[244]. Some of the

recently described Cl-amidine analogues, however, such as BB-Cl-amidine, have demonstrated higher degrees of PAD inhibition[246]. BB-Cl-amidine is a Cl-amidine derivative retaining the function of the original molecule but possessing a C-terminal benzimidazole used to restrict proteolysis of the C-terminal amide. This resulted in a 20-fold increase in the cellular potency of BB-Cl-amidine against PAD4 as compared to Cl-amidine, whilst simultaneously increasing its *in vivo* half life[246].

Reversible PAD inhibitors include the chemotherapy agent paclitaxel, Bz-N^G-monomethyl-Arg and benzoyl-N^G,N^G-dimethyl-Arg (Bz-ADMA)[247,248], however, these compounds are much less potent than Cl-amidine and therefore are not used to inhibit PADs. In conclusion, BB-Cl-amidine is currently the best recognised inhibitor of PADs. Nonetheless, although it has been widely adopted for studies of citrullination *in vivo* and *in vitro*, it has a number of limitations, including a lack of specificity to certain PADs, relatively low potency, limited bioavailability, and a short *in vivo* half-life. Due to this, further research should focus on developing or identifying inhibitors that overcome these limitations.

CHAPTER 2. MATERIALS AND METHODS

2.1. Human samples

Human hepatic metastasis tissues with surrounding unaffected liver tissues, as well as primary colon adenocarcinoma tissues with adjacent unaffected colon tissues, were obtained from the Oxford Radcliffe Biobank following institutional review and granting of ethical approval. All patient data were anonymised and had no clinical data available. All bioinformatics data were anonymised and required no ethical approval.

2.2. Cell lines

The following cell lines were used in the study: SW480, SW620, DLD1, HT29, HCT116, LoVo (all human colon cancer), MC38 (mouse colon cancer), OE33 (human oesophageal cancer), HeLa (human cervical cancer) and 2H-11 (mouse endothelial cell line). All cell lines were purchased from American Type Culture Collection (ATCC). Early passage SW480, SW620, DLD1, HT29, HCT116 and OE33 cell lines were cultured in Roswell Park Memorial Institute (RPMI-1640) media (R8758; Sigma). Early passage MC38, HeLa, and 2H-11 cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM) media (D2429; Sigma). All media were supplemented with 10% foetal bovine serum (FBS, 16000-044; Fisher), and 100 IU/ml penicillin/streptomycin (15140-122; Fisher). All cells were cultured under sterile conditions and maintained in a 37°C incubator with 5% CO_2, regularly tested for mycoplasma and genetically authenticated by CRUK Cancer Centre Genomics Facility, Leeds, UK. Cell lines were passaged upon reaching approximately 80% confluency and their morphology was regularly checked to ensure the absence of cross contamination between cell lines. To determine cell number, the automated cell counter NucleoCounter NC-100 (Chemometec) was used.

For experiments presented in Figure 20, torkinib was purchased from Selleckchem (S2218). It is a selective mTOR inhibitor with IC50 of 8 nM in cell-free assays; targets both mTOR complexes with >10- and 100-fold selectivity for mTOR than PI3Kδ or PI3Kα/β/γ, respectively. Torkinib was used at 2.5 nM.

2.3. Cell transfection

For the generation of cell lines stably expressing green fluorescent protein (GFP), cells were infected with a lentivirus containing a gene encoding GFP driven by cytomegalovirus (CMV) promoter (CLS-PCG-8; Qiagen). For transfection, 5×10^3 cancer cells were spun at 3000 rpm. After that, cells were kept for 30 min at 4°C in media supplemented with 8 µg/ml polybrene (H9268; Sigma-Aldrich) and lentivirus particles at various multiplicities of infection (2-8). After spinoculation, cells were plated in 12-well plates and cultured in a 37°C incubator with 5% CO_2. Two days post transfection, lentivirus-containing media was replaced with fresh media containing puromycin (P8833; Sigma-Aldrich) at a pre-determined concentration (usually 2-5 µg/ml). Several weeks after, 10% of the top GFP-positive cells were selected by fluorescence activated cell sorting (FACS) in the Jenner Institute with the help of Dr. Andrew Worth. Following sorting, cells were cultured in puromycin-containing media at a pre-determined concentration.

For generation of shPAD4 cells, the PADI4 Human shRNA Lentiviral Particle (Locus ID 23569) kit was purchased (TL310630V, OriGene). Lentivirus contained puromycin resistance cassette for selection of transfected cells and GFP gene for detection. Control cells were transfected with lentivirus particles harbouring scramble shRNA. The map of shRNA cloning vector is available at https://www.origene.com/catalog/rnai/shrna-lentiviral-particle/tl310630v/padi4-human-

shrna-lentiviral-particle-locus-id-23569. Transfected cells were used as a pool for selection rather than established individual colonies.

2.4. Animal experiments

All animal experiments were performed according to the UK Animal Scientific Procedures Act 1986, institutional guidelines of the University of Oxford, as well as published guidelines for the welfare and use of animals in cancer research[249]. All experiments involving mice were conducted within the limits of the Project License issued by the Home Office, United Kingdom (PPL numbers 30/2841 and 30/3413). Animals were housed in specific pathogen-free facilities with controlled conditions of temperature (24±1°C) and humidity (55±10%) and had access to water and food (rodent pellets) *ad libitum.* Throughout the whole time of experiment 12 h light/dark cycles were carefully maintained. All mice were weighed twice weekly and their health condition was monitored every day by myself or Biomedical Science facility personnel.

Female severe combined immunodeficiency (SCID) mice aged 6-8 weeks were purchased from Harlan Laboratory (Indianapolis, IN). Female C57BL/6 mice aged 6-8 weeks were purchased from Charles River Laboratories (Kent, UK).

At the end point of each experiment, mice were humanely culled by means of intraperitoneal injection of 200 µl pentobarbitone, and livers, subcutaneous tumours or caecal tumours were dissected thereafter.

For hepatic metastases model, mice were first anesthetised using vapourised isoflurane. After that, the upper lateral abdominal wall was incised and 1×10^6 HT29,

HCT116, or LoVo cells, or 5×10^5 MC38 cells in 100 µl phosphate-buffered saline (PBS) were injected into parenchyma of the spleen. The spleen was removed by electrocautery after 1 min post injection to prevent primary tumour growth in the splenic bed. The wound was then closed using non-soluble sutures. SCID mice injected with HT29, HCT116 or LoVo cells were sacrificed 35±2 days following surgery. C57BL/6 mice injected with MC38 cells were sacrificed 13±2 days following surgery. For subcutaneous tumour model, 1×10^6 HT29, HCT116, LoVo, or MC38 cancer cells were injected into the left flank of anesthetised C57BL/6 or SCID mice. Tumour volume (in mm^3) was determined every other day by calliper measurement of the length, width, and height of each tumour. For intercaecal injection, mice were first anesthetised using vaporised isoflurane. After that, a transverse upper abdominal wall incision was made, the cecum was exteriorised, and 1×10^6 wild-type or shPAD4 HCT116 cells in 50 µl PBS were slowly injected between the mucosa and the muscularis layers of the caecal wall. After injection, the cecum was returned to the abdominal cavity and the wound was closed with surgical sutures and autoclips. For *in vivo* PAD4 inhibition experiments, randomised SCID or C57BL/6 mice were administered with 7.5 mg/kg BB-Cl-amidine or 100 µl PBS via the intraperitoneal route every 48 h starting at day 1 after intrasplenic injection. The mice were weighed daily in order to monitor toxic effects of the drug.

2.5. BB-Cl-amidine

The BB-Cl-amidine compound was synthesised by Prof. Paul Thompson`s lab as described[246]. This compound retains the critical elements of Cl-amidine but possesses a C-terminal benzimidazole designed to limit proteolysis of the C-terminal amide. This greatly facilitates cellular uptake of the drug due to increase in hydrophobicity. Lyophilised BB-Cl-amidine was diluted in 100% dimethyl sulfoxide (DMSO), aliquot

and stored at -20°C before use. For *in vivo* administration, 3 µl BB-Cl-amidine was diluted in 97 µl PBS (1:100) to reach the desired concentration of 7.5 mg/kg.

2.6. Tissue decellularisation

For liver tissue decellularisation, a solution of 0.1% sodium dodecyl sulphate (SDS, L3771; Sigma) and 0.01% ammonium hydroxide (NH_4OH, 221228; Fluka) was used. This formulation was previously identified to produce the fastest and most thorough decellularisation[250]. Freshly resected specimens of liver metastases and surrounding unaffected livers were delivered from the hospital to the laboratory in RPMI media on ice and then placed into petri dishes filled with the decellularising solution. The dishes were put on a rotary shaker for 72 h; the buffer was changed every 6 h.

2.7. Initial ECM preparation

Decellularised tissues were cut in 100 mg chunks (wet weight), aliquot into 200 µL of ice cold buffer C (N-2- hydroxyethylpiperazine-N- 2-ethane sulfonic acid [HEPES] pH 7.9, $MgCl_2$, KCl, ethylenediaminetetraacetic acid (EDTA), Sucrose, Glycerol, Sodium OrthoVanadate, protease inhibitor cocktail, concentration of each component is not specified by a supplier due to trade secret) of the CNMCS Compartmental Protein Extraction Kit (K3013010; BioChain Institute), snap frozen and stored at -30°C until use.

2.8. ECM enrichment

ECM samples were thawed on ice, then homogenised using the electric blade tissue homogeniser (985370-07; BioSpec Products) at speed 4-5 for 20-25 s. Homogenisation was repeated three times and the probe was cleaned with ethanol

between samples. The samples were further processed on high intensity ultrasonic processor at 4°C at 100% amplitude and the following pulse mode: 2 s on/2 s off, 20 s in total (including offs). In order to remove complex oligosaccharides from glycosylated proteins and also eliminate contaminating nucleic acids, the samples were then treated with 5000 units of PNGase F (P0704; New England BioLabs) and 1 μL Benzonase® nuclease (E1014; Sigma) for 1 h at 4°C. The samples were spun at 18,000 x g for 15 min, before the supernatant was discarded, and the pellet was washed in 400 μL of ice cold buffer W (HEPES (pH 7.9), MgCl2, KCl, EDTA, Sucrose, Glycerol, and Sodium OrthoVanadate, CNMCS Compartmental Protein Extraction Kit (K3013010; BioChain Institute), concentration of each component is not specified by a supplier due to trade secret) at 4˚C for 5 min. The resulting washed protein extracts were spun at 18,000 x g for 15 min. The supernatants were then discarded, and the pellets were resuspended in 150 μL of ice cold buffer N (HEPES (pH 7.9), $MgCl_2$, NaCl, EDTA, Glycerol, Sodium OrthoVanadate, CNMCS Compartmental Protein Extraction Kit (K3013010; BioChain Institute), concentration of each component is not specified by a supplier due to trade secret) and incubated at 4°C for 20 min to solubilise nuclear proteins. The protein extracts were spun at 18,000 x g for 20 min. Supernatants were withdrawn, and the pellets were resuspended in 150 μL of ice cold buffer M (HEPES (pH 7.9), MgCl2, KCl, EDTA, Sucrose, Glycerol, Sodium deoxycholate, NP-40, Sodium OrthoVanadate, CNMCS Compartmental Protein Extraction Kit (K3013010; BioChain Institute), concentration of each component is not specified by a supplier due to trade secret) for solubilisation of membrane-bound proteins. The samples were spun 18,000 x g for 15 min and supernatants were discarded. The resulting pellets were resuspended in 150 μL of prewarmed buffer CS (Pipes (pH 6.8), $MgCl_2$, NaCl, EDTA, Sucrose, SDS, and Sodium OrthoVanadate, CNMCS Compartmental Protein Extraction Kit (K3013010; BioChain

Institute), concentration of each component is not specified by a supplier due to trade secret) and incubated at room temperature for 20 min to solubilise cytoskeleton proteins. The extracts were spun at 18,000 x g for 20 min. Supernatants were discarded, and the pellet was finally resuspended in 150 μL of buffer C, incubated at 4°C for 5 min and then spun for 20 min at 15,000 x rpm at 4°C. The remaining insoluble pellets of ECM proteins were snap frozen and stored at -30°C until use.

2.9. Solubilisation of ECM-enriched proteins

In order to solubilise ECM-enriched extracts, the samples were thawed on ice and then a cocktail of reducing agents, namely, 8M urea, 100 mM ammonium bicarbonate, 10 mM dithiothreitol (DTT, 43817; Sigma), pH 7.8, was added to each sample. Samples were then incubated at 37°C for 30-60 min. Alkylating agent iodoacetamide (I1149; Sigma) was added to a final concentration of 25 mM and the samples were incubated for 30 min at room temperature. At this stage, protein concentration was measured using the detergent-compatible protein assay (5000111; Bio-Rad) according to manufacturer`s instructions. Protein samples were then precipitated by the methanol/chloroform extraction method, followed by resuspension in 50 μL of 8M urea and vortexing. Finally, the urea concentration was reduced to a final concentration of 1M by diluting the reaction mixture with 250 μL of ddH_2O. Finally, ECM-enriched pellets were digested overnight using trypsin (V5111; Promega), at a ratio of 1:50 enzyme:substrate. The samples were constantly vortexed at 37°C. After that, a second aliquot of trypsin was added at a ratio of 1:100 enzyme:substrate, and samples were incubated for an additional 4-6 h.

2.10. Mass-spectrometry

The samples were processed with nano–liquid chromatography tandem mass spectrometry (nano–LC-MS/MS), using the Acquity LC instrument (C18 column with a 75 µmx 250mm, 1.7 µm particle size; Nanoacquity Waters) coupled to a Thermo LTQ Orbitrap Elite mass spectrometer (resolution of 120,000 at 400 m/z, Top 20, collision-induced dissociation), using a gradient of 1–35% acetonitrile for 60 min at a flow rate of 250 nl/min. Peak lists containing MS/MS spectra were generated using MSConvert (Proteome Wizard). Following this, peak lists were searched utilising Mascot version 2.3 (http://www.matrixscience.com), against the Swiss-Prot protein database containing mouse (16,642 entries as of September 2012) or human (20,306 entries as of June 2014) sequences, with tryptic restriction and mass deviations of 10 parts per million/0.5 Daltons in the respective MS modes. Oxidation of methionine, deamidation of asparagine and glutamine, and other known collagen and proteoglycan modifications were used as variable modifications. No differential posttranslational modifications were used. For label-free quantification of differentially expressed proteins, normalised abundance of each protein was determined by measuring the peak area intensity utilizing the Progenesis QI software (Nonlinear Dynamics). Protein abundance was calculated from the sum of all unique peptide ion abundances for a specific protein on each run. Normalisation of abundance was performed to allow comparisons across different sample runs by the software. Proteins identified by more than one peptide were retained. The normalised peptide intensities for each sample were used to calculate fold change ratios for proteins between groups.

2.11. Protein categorisation

In order to pinpoint ECM proteins out of the entire array of LC-MS/MS hits, a categorisation proposed by Naba et al.[251] was used. The ensemble of ECM proteins was split into five major groups: proteoglycans, glycoproteins, collagens, ECM regulators, ECM-affiliated proteins and secreted factors. For the experiments shown in Figure 8, a gene ontology (GO) term enrichment analysis for cellular localisation was performed using the PANTHER bioinformatics resource (http://pantherdb.org/)[252] in order to determine whether these hits belonged to the ECM.

2.12. LC-MS/MS detection of citrulline residues

LC-MS/MS data were analysed in Peaks software (V.6; Bioinformatics Solutions, Waterloo, ON, Canada) using the post-translational modification search option at mass tolerances of 10 ppm for MS1 and 0.05 Da for MS2 at 1% false discovery rate (FDR). Citrullinated peptides were manually inspected when the citrullinated residue produced a missed cleavage to exclude false positive assignment of citrullination (that is deamidation or the selection of the first isotopic peak as precursor mass).

2.13. Colony growth assay

HT29, HCT116, LoVo, MC38, and 2H-11 cells (100-300 cells/well) were plated in 6-well plates and allowed to adhere to culture dishes for 4 h. A range of BB-Cl-amidine concentrations (0.05-5 μM) in 2 ml of RPMI 1640 was added to culture dishes. Colonies were stained with 1% crystal violet in ethanol and counted using a colony counter (GelCount Tumour Colony Counter, Oxford Optronix) 10 to 20 d after treatment. Survival was expressed as the relative plating efficiencies of the treated cells compared

with that of the DMSO-treated (i.e., control) cells. Experiments were performed at least 2 times with 6 replicates per condition.

2.14. mRNA extraction and qPCR

For real-time polymerase chain reaction (PCR) analysis, total RNA was extracted using RNeasy Mini Kit (74104; Qiagen) from FACS-sorted liver metastasis cells or cancer cells collected from culture dishes, according to manufacturer`s instructions. RNA concentration was determined using NanoDrop 2000 (Thermo Fisher Scientific). Total RNA was reverse-transcribed into cDNA using the Tetro cDNA Synthesis Kit (BIO-65043; Bioline) following the manufacturer`s protocol. Contaminating DNA was eliminated using the TURBO DNA-free™ kit (AM1907; Ambion) according to manufacturer`s instructions. PCR for PAD4, and hypoxanthine phosphoribosyltransferase (HPRT) was performed on 250 ng of cDNA using SybrGreen technology and the StrataGene Mx3005P QPCR System (Aligent) according to manufacturer`s instructions. For analysis, ΔCt method was used and fold change was calculated ($2^{-\Delta\Delta Ct}$). All results were normalised to HPRT expression. The following real-time PCR primers were used: human PAD4 forward (5′-TTGCAATCAACTGGAGCAGG- 3′) and reverse (5′- AGAGGGCTGAGGCCACCT-3′), mouse PAD4 forward (5′-TTACCACGGAGTTCCACACC- 3′) and reverse (5′-CTGCCAGCCATAGTTAAGC- 3′), human HPRT forward (5′-ATAGGACTCCAGATGTTTCC-3′) and reverse (5′-ATAAGCCAGACTTTGTTGG-3′), mouse HPRT forward (5′-GCAGTACAGCCCCAAAATGG- 3′) and reverse (5′-AACAAAGTCTGGCCTGTATCCAA-3′). All other primers used in the study were validated and optimised by manufacturer and purchased from Sigma.

For RT² Profiler™ PCR arrays (PAHS-090Z and PAMM-090Z; Qiagen), total RNA was reverse-transcribed to cDNA using the First Strand cDNA Synthesis kit (330401; Qiagen) following the manufacturer`s protocol. RT² qPCR Mastermix (330520; Qiagen) was used for PCR reaction according to manufacturer`s instructions.

2.15. Isolation of human neutrophils and monocytes

Ficoll-Paque™ density gradient media was used to isolate neutrophils and monocytes from a 24 year old healthy male volunteer, following extraction of total protein, determination of its concentration and storage at -30ºC. Briefly, blood was diluted 2-4x the volume of a buffer containing PBS (pH 7.2) and 2 mM EDTA, and then 35 mL of diluted cell suspension was layered over 15 mL of Ficoll-Paque™ in a 50 mL conical tube. After this, the tube was centrifuged at 400×g for 30-40 min at RT. After centrifugation, layers of mononuclear cells and granulocytes were carefully isolated and lysed using RIPA buffer followed by protein extraction.

2.16. Exosome isolation, characterisation and analysis

HT29, HCT116 and LoVo cells were plated at high density (1.5×10^7) into T175 culture flasks. To produce exosome-rich conditioned media, cells were washed three times with PBS and incubated for 72 h with DMEM supplemented with 0.1% FBS depleted of bovine exosomes by overnight centrifugation at 100000×g. The resulting conditioned media was collected and sequentially centrifuged at 2000×g for 30 min, 10000×g for 40 min, and 100000×g for 18 h. After each centrifugation, the pellets containing cell debris were discarded and supernatant was used for the next step. Resulting pellets were finally washed in ice-cold PBS and further centrifuged at

100000×g for 1.5 h. All centrifugation steps were performed at 4ºC. To verify exosome preparations, collected exosomes were resuspended in 15 µL PBS and then processed for transmission electron microscopy analysis in the Sir William Dunn School of Pathology, University of Oxford. In brief, exosome suspension was fixed in 100 µL of 2% paraformaldehyde (PFA). This mix was then transferred onto carbon coated electron microscopy grids. Membranes were covered for 20 min and then washed with 1% glutaraldehyde (5 min) and distilled water. No contrasting agent was used during microscopy.

For western blotting analysis of exosomes, final exosome-containing pellets were resuspended in 40 µL RIPA lysis buffer and protein concentration was determined using micro BSA assay kit (23235; Thermo Fisher) according to manufacturer`s instructions.

To quantify exosomes in solution and determine their size, the nanoparticle tracking analysis (NTA) analysis was performed using the NanoSight NS300 (Malvern) instrument.

2.17. Cell attachment assays

For real-time cell attachment analyses, the xCELLigence RTCA System was used (ACEA Biosciences) according to manufacturer`s protocol. Briefly, gold microelectrode E-plates were coated overnight with full length Collagen I alone or together with active human PAD4 or 75 µM BB-Cl-amidine. Wells were washed 3 times with PBS, and 15×10^3 cells were plated in triplicates per condition and analysed on the xCELLigence system with impedance readings taken every 15 min to 18 h. Impedance measurements were used to calculate cell index which is proportional to the number of cells adhered.

For classical adhesion assay, 5×10^4 GFP^+ cells were seeded in triplicates in 6-well plates precoated overnight with collagen I, collagen I+PAD4, or collagen I+PAD4+BB-Cl-amidine (undiluted). For integrin blocking, the cells were pre-incubated with integrin blocking antibodies for 2 h at concentrations provided by supplier before seeding. At 1, 2 and 3 h time points, the plates were imaged using an inverted epifluorescence microscope at 10x objective (DM IRBE, Leica Microsystems, at least 5 fields of view per condition) and the number of adherent cells was manually counted.

2.18. GW4869

GW4869 was purchased from Sigma (D1692). Cancer cells were plated at high density (10×10^6) and serum starved for 8 h. Media was changed for Opti-MEM™ (31985070; Thermo Fisher) and 20 μM GW4869 was added to cells at time points 0, 24 and 48 h. The serum was collected at 72 h and exosomes were isolated as described above.

2.19. Time-lapse imaging and cell tracking analysis

Cancer cells stably expressing GFP were plated at low density (0.8×10^3) in 96-well plates, 8 or 16 wells per condition. Prior to analysis, the wells were coated overnight with full length Collagen I alone or together with active human PAD4 or 100 μM BB-Cl- amidine. Cells were imaged every 1 h for a total of 12-24 h using a time-lapse Nikon Ti-E microscope. Resulting videos were merged and processed with IMARIS software (Bitplane) to quantitatively analyse cell displacement length and median speed.

2.20. Survival analysis

Survival analysis was performed using GraphPad Prism for colorectal cancer data derived from cBioportal[253] .We used JetSet probes throughout[254] and patients were divided into two groups on the basis of median expression of the 13 gene signature. For the analysis of colorectal cancer datasets, a z-score threshold of þ1 for gene expression was used to define patients. Kaplan–Meier survival curves were constructed and compared using the log-rank method to generate hazard ratios and P values. Survival curves were generated using GraphPad Prism Version 7.

2.21. Antibodies

The list of antibodies used in the study is provided in the Table 2.

Table 2. List of antibodies used in the study

Antibody	Company	Catalogue number	Application
Laminin	Abcam	ab30320	IHC
Collagen V	Abcam	ab7046	IHC
Collagen IV	Abcam	ab6586	IHC
Collagen IV	Abcam	ab23975	IHC
MMP-23	Abcam	ab53148	IHC
COL5A1	Santa-Cruz	sc-20648	IHC
PAD4	Acris	AP55150SU-N	IHC
Collagen I	Abcam	ab90395	IHC
N-cadherin	Cell Signaling	13116	IHC
Cytokeratin 7	Cell Signaling	4465	IHC

Fibronectin	Abcam	ab2413	IHC
Anti-Histone H3 (citrulline R2 + R8 + R17)	Abcam	ab5103	IHC
Goat Anti-Rabbit IgG H&L (Alexa Fluor® 488)	Abcam	ab150077	IHC
Goat Anti-Mouse IgG H&L (Alexa Fluor® 488)	Abcam	ab150113	IHC
Donkey Anti-Rabbit IgG H&L (Alexa Fluor® 555)	Abcam	ab150074	IHC
Donkey Anti-Mouse IgG H&L (Alexa Fluor® 555)	Abcam	ab150106	IHC
PAD4	Abcam	ab128086	WB
PAD4	Abcam	ab26071	WB
Col5A1	Santa-Cruz	sc-20648	WB
Versican	Santa-Cruz	sc-25831	WB
Tenascin C	Abcam	ab108930	WB
MMP-23	Abcam	ab53148	WB
S100-A11	ProteinTech	10237-1- AP	WB
Caveolin-1	Cell Signaling	3238	WB
B-catenin	Santa-Cruz	sc-7199	WB
Histone H3	Cell Signaling	4499	WB
Zeb-1	Cell Signaling	3396	WB
Cytokeratin 7	Cell Signaling	4465	WB
N-cadherin	Cell Signaling	13116	WB
E-cadherin	Cell Signaling	5296	WB

N-cadherin	Cell Signaling	4061	WB
E-cadherin	R&D	AF748	WB
ZO-1	Cell Signaling	8193	WB
MEK1/2	Cell Signaling	9126	WB
p44/42 MAPK	Cell Signaling	4377	WB
JNK	Cell Signaling	9252	WB
pJNK	Cell Signaling	9251	WB
FAK	Cell Signaling	13009	WB
pFAK	Cell Signaling	8556	WB
B-actin	Cell Signaling	8457	WB
GAPDH	Cell Signaling	5174	WB
Citrullinated residues	Millipore	17-347	WB
Donkey anti-goat IgG-HRP	Santa-Cruz	sc-2020	WB
Goat anti-rabbit IgG-HRP	Santa-Cruz	sc-2030	WB
Goat anti-mouse IgG-HRP	Santa-Cruz	sc-2031	WB
Anti-Integrin alpha 2+β 1	Abcam	ab24697	Blocking
Anti-Integrin α1	Millipore	MAB1973	Blocking
CD49a PerCP-Cy™5.5	BD Biosciences	564862	Blocking
Integrin alpha 2 β 1	Novus Biologicals	NBP2-29787	Blocking
CD11b PE-Cy7	eBioscience	25-0112-82	FACS
CD45 PE	eBioscience	12-0451-82	FACS
CXCR2-APC	Biolegend	149305	FACS

Gr1 PE	eBioscience	12-5931-82	FACS
CD45 PE-Cy7	Biolegend	103113	FACS
CD140a APC	Biolegend	135907	FACS
CD31 PE	Biolegend	102417	FACS

2.22. *In vitro* citrullination studies

For all experiments involving *in vitro* citrullination, the citrullination assays were performed as previously described[249]. Briefly, recombinant collagen I (ab6308; Abcam) was incubated at a concentration of 40 μg/mL with 200 ng active human PAD4 (ab196393; Abcam) in the presence or absence of 1 mM or 0.05 mM BB-Cl-amidine for 18 h at 37ºC in a buffer of 20 mM HEPES (pH 8.8), 10 mM $CaCl_2$, 1 mM DTT, 1 mM EDTA, and 0.3 M NaCl. The concentrations of PAD4 and collagen I were used to obtain the enzyme-to-substrate ratio of 1:10, as previously described[255]. The reaction was stopped by the addition of EDTA to a final concentration of 10 mM. Samples were analysed by LC-MS/MS for the presence of citrulline residues. For the coating experiments, the mixture of collagen I and PAD4 (and BB-Cl-amidine if needed) was immediately added into wells (96-well, 24-well or 6-well) and then incubated for 18 h at 37ºC. Before plating the cells, the wells were dried in a 37ºC incubator and carefully washed with PBS.

2.23. Immunofluorescence and H&E staining

Liver metastasis tissues, adjacent unaffected liver tissues, and murine- or human-derived decellularised matrices were embedded in Optimal Cutting Temperature (OCT) compound (25608-930; Tissue-Tek), flash frozen and kept at -80ºC until cutting. Tissue

sections were prepared using the OTF 5000 cryostat (Bright instrument), and 12 μm slices were mounted on Superfrost Plus microscope slides (4951PLUS4; Thermo Fisher). For immunofluorescence, the sections were brought to RT, fixed in ice-cold acetone, rehydrated, blocked with 20% goat serum, and incubated overnight with the primary antibodies. The following day, the slides were washed in PBS before being incubated with secondary antibodies for 1 h at room temperature (RT). The sections were washed and mounted using the ProLong® Diamond Antifade Mountant with 4',6-diamidino-2-phenylindole (DAPI) (P36962; Thermo Fisher). Immunofluorescence was visualised utilizing an inverted epifluorescence microscope (DM IRBE, Leica Microsystems) with a digital camera. All images were acquired with either a 10x or a 20x objective. The excitation source was CoolED pE-300 ultra. The images were processed using either ImageJ (National Institutes of Health, Bethesda, MD) or Adobe Photoshop CS4 (Adobe Systems). For H&E staining, slides were fixed in ice-cold acetone for 10 min, then immersed in filtered Harris modified Haematoxylin solution (HHS16; Sigma) for 1 min, then rinsed with tap water. The sections were immersed in 1% aqueous Eosin Y solution (HT110216; Sigma) for 1-2 min and rinsed with tap water. The slides were dehydrated in ascending alcohol solutions (50%, 70%, 80%, 95% x 2, and 100% x 2) and cleared with xylene two times for 5 min each. Finally, coverslips were mounted onto a labelled glass slide with Permount Mounting Medium (SP15-100; Thermo Fisher). For quantification of liver metastatic areas, H&E-stained sections of livers were imaged using digital slide scanner (Leica) and percent of metastatic area in was quantified using the ImageJ software.

2.24. Protein extraction from cells and tissues

For protein extraction from cells, cell pellets were lysed with RIPA lysis buffer (89900; Thermo Fisher), containing a Complete Protease Inhibitor Cocktail (04693116001; Roche) at a ratio of 5:1, for 30 min at 4ºC on a rotary shaker. For protein extraction from intact or decellularised tissues, 0.5 g of freshly dissected tissue was disrupted at 4 °C using an electric Tissue-Tearor homogeniser, passed through a 70 μm cell strainer and spun down. Cell pellets were resuspended in RIPA lysis buffer containing a cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail (04693159001; Sigma) at a ratio of 5:1 for 2 h at 4ºC on a rotary shaker. The extracts were briefly sonicated and then spun down at 14,000 x g at 4ºC for 10 min to remove insoluble material. Protein concentration was determined using the BCA protein assay kit (23225; Thermo Fisher). All protein suspensions were flash frozen in liquid nitrogen and kept at -20ºC until use.

2.25. Immunoblot analysis

For immunoblotting, an equal amount of protein lysates (8-60 μg, depending on the primary antibody used) were mixed with NuPAGE™ LDS Sample Buffer (NP0007; Thermo Fisher) at a ratio 4:1, and loaded on either 3-8% Tris-Acetate or 4-12% Bis-Tris precast gels (NP0321; Life Technologies) for high and low molecular weight proteins, respectively. Proteins were separated at 100 V for 1.5-3 h and transferred at 30 V for 1.5 h 4ºC on polyvinylidene fluoride membrane (IPVH00010; Millipore). Membranes were then blocked in either 5% bovine serum albumin (A2153; Sigma) or 5% skimmed milk diluted in 100% tris-based saline (TBS) in 0.05% Tween (TBST). Blots were incubated overnight at 4ºC with the primary antibodies at dilution 1:1000. The membranes were washed 3 times for 10 min each in TBST, then horseradish peroxidase conjugated

secondary antibodies were added at dilution 1:1000 for 1 h at RT. Proteins were visualised with enhanced chemiluminescence by using the Amersham ECL detection reagent (RPN2106; GE Healthcare) and either the Odyssey® imaging system (LI-COR) or an X-ray detector. Densitometric analysis of band intensity was performed using the ImageJ software or the Image Studio™ Software (LI-COR).

For densitometry analysis, pixel density for a protein of interest was normalised to pixel density of loading control.

For analysis presented in Figure 19, the Proteome Profiler Human Phospho-Kinase Array Kit was used (ARY003B; R&D) according to manufacturer`s instructions (some of the proteins were non-phosphorylated only).

2.26. Citrullination immunoblot analysis

For determination of citrulline levels in liver metastases tissues and adjacent unaffected liver tissues, a modified anti-citrulline detection kit (17-347; Merck Millipore) was used according to manufacturer`s instructions. The assay principle is based on immunoblot technique using the anti-citrulline antibody. To detect citrulline residues, the polyvinylidene fluoride membrane with transferred proteins was exposed overnight to 2,3-butanedione monoxime and antipyrine in a strong acid solution to form uriedo group adducts to citrulline residues. This ensures specific detection of citrulline containing proteins disregarding any other amino acid sequences.

2.27. ELISA

For determination of citrulline levels in human tissues, an anti-citrulline human ELISA kit (MBS701429, MyBioSource) was used according to manufacturer`s instructions. For determination of PAD4 levels in human tissues or in conditioned media

from cells, an anti-PAD4 human ELISA kit (LS-F7962, LifeSpan BioSciences or MBS2505910, MyBioSource) was used according to manufacturer`s instructions. All ELISA were performed with 10 μg of protein.

2.28. SDS-PAGE and gel silver staining

For sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and gel silver staining, total protein (5 μg) obtained from intact tissue, decellularised tissue, or decellularised biochemically enriched tissue, were mixed with LDS sample buffer (NP0008, Life Technologies) at a ratio 4:1, and then loaded on 1.0 mm 3-8% Tris-Acetate protein gradient gel (EA0375, Life Technologies). Proteins were separated at 100 V for 2 h. Silver staining of gels was performed using the SilverQuest staining kit (LC6070; Life Technologies) according to the manufacturer's instructions. Images of the stained gels were acquired with the PhotoDoc-It UVP Imaging System (Cole-Parmer, UK).

2.29. WST-1 proliferation assay

5×10^3 cells were seeded in 96-well plates. 5 technical replicates per condition were used. After 24 h of culture, media was replaced with fresh media containing 10% WST-1 reagent and cells were left in the incubator for 1 h. Absorbance was read at 450 nm using the plate reader.

2.30. Cell sorting

Dissected mouse metastatic livers were minced on ice and incubated at 37˚C in RPMI-1640 media containing 0.05% collagenase/dispase (11097113001; Roche) and

0.01% trypsin inhibitor (T7659; Sigma). The lysate was filtered through a 70 μm cell strainer into PBS and residual red blood cells lysed using red cell lysis buffer (11814389001; Roche). Cell suspensions at a concentration of 1×10^7 cells/ml were the incubated for 45 min at RT in the presence of anti-Gr1-PE (RB6-8C5; eBioscience), anti-CD140a-APC (135907; Biolegend), anti-CD31-PE (102507; Biolegend), anti-CD45-PE-Cy7 (Biolegend; 103113), anti-CD11b-PE-Cy7 (25-0112-81; eBioscience), and Fc blocking antibody (14-9161-71, eBioscience), before being sorted at Legacy MoFlo MLS High Speed Cell Sorter (Beckman Coulter). Tumour cells were considered GFP^+ cells, fibroblasts were considered $CD140a^+$, endothelial cells were considered CD31+ CD45-, immune cells were considered $CD45^+$ $CD31^-$, neutrophils were considered $CD11b^+$ $Gr1^+$ cells, myeloid cells were considered $CD11b^+$ $Gr1^+$ cells, other stromal cells were considered GFP^- $CD11b^-$ $Gr1^-$ cells. For flow plot analyses, samples were run using the FlowJo 7.5 software (Tree Star).

2.31. Neutrophil depletion

For neutrophil depletion experiments, 12.5 μg/kg of anti-Ly6G (clone 1A8; BD Biosciences, Oxford, UK) was injected intraperitoneally every 24 h. An isotype control antibody immunoglobulin [Ig] G2a was injected as a control.

2.32. Statistical analysis

Data were analysed, and statistics were performed in Graph Pad Prism 5. Data are presented as mean±s.e.m or median±interquartile range where indicated. Statistical analyses were carried out using the paired Student's t-test. Values with $P < 0.05$ were considered statistically significant. FDR was used to validate peptide and protein hits obtained during the quantitative LC-MS/MS analysis. Two-way analysis of variance

(ANOVA) was used to identify the difference between groups in the quantitative proteomics analysis, subcutaneous tumour growth curves, and WST-1 proliferation assays. Kruskal-Wallis test with Dunn`s multiple comparison post-test was used for comparison of time-lapse microscopy data. In all other analyses, two-tailed paired or unpaired Student's t-test was used. Statistical methods were not used to predetermine necessary sample size, but sample sizes were based on review of similar experiments in literature and previously published results such that appropriate statistical tests could yield significant results. Experiments were not performed in a blinded fashion. The following symbols represent significant differences throughout the figures: *P = 0.01 to 0.05, **P = 0.001 to 0.01, ***P < 0.0001.

CHAPTER 3. QUANTITATIVE PROTEOMICS IDENTIFIES ECM PROTEINS PREDICTIVE OF SURVIVAL IN COLON CANCER PATIENTS

3.1. Introduction

3.1.1. Defining the matrisome

During the last several decades, there has been a growing interest to the field of the ECM, especially in the context of human disease[256]. However, ECM research suffered from the lack of understanding of the hierarchy, nomenclature, and classification of ECM proteins, limiting systematic investigation of this major group of molecules. This problem was addressed several years ago, when a research group by Richard Hynes from the Massachusetts Institute of Technology developed a complicated computer-aided approach to categorise the ensemble of all existing ECM and ECM-associated proteins in human and mouse[251]. This bioinformatics method utilised an *in silico* tool enabling to predict all ECM proteins on the basis of 55 conserved protein domains characteristic of these molecules. This strategy enabled Hynes and colleagues to comprehensively characterise the human matrisome, thus creating an opportunity for systematic research in the field. In his classification, Hynes selected the core matrisome, consisting of three major groups of proteins, namely, collagens, glycoproteins, and proteoglycans[251]. Apart from the core matrisome, Hynes also selected ECM regulators - these include various protease inhibitors, cross-linkers, ECM-remodelling enzymes and proteases, which can either interact with or bind to ECM proteins in order to modify their structure and/or function[251]. Alterations in expression or activity of many of these factors have been identified during cancer progression where they are related to many of the fundamental

hallmarks of cancer including invasion, angiogenesis and hypoxia[257], indicating a crucial link between the biomechanical properties of the ECM and cancer progression.

As well as these classes of proteins, he introduced a group of secreted factors including cytokines, growth factors and chemokines that are able to interact with ECM proteins are important for determining the functional properties of the ECM. Finally, the last group of ECM-affiliated proteins comprised the rest of the factors structurally or physically affiliated with the core matrisomal proteins (this includes mucins, annexins, galectins, etc). The results presented in this thesis were classified and categorised in accordance with the matrisome dataset as the only available system defining the whole ensemble of ECM molecules.

3.1.1. Collagens

Fibrillar collagens (I-III, V, XI, XXIV and XXVII) represent the most abundant group of proteins in the human body. They account for around 35% of the total body protein, being the main structural component of the ligament, tendon, bone, cartilage and skin[258]. Collagens are also important constituents of blood vessels, teeth, and the cornea. Although collagen type I is the most abundant collagen in mammals (up to 90%), there are a total of 28 types of collagens, expression of which is cell- and tissue-specific[258]. Structurally, collagens are formed by multiple post-translational modifications of the newly synthesised collagen pre-peptide, followed by the formation of pro-collagen and it's Golgi-mediated transport into the ECM[259]. In the extracellular space, pro-collagen molecules are self-assembled into triple helical fibres usually 10 to 300 nm in diameter[259]. At the final step of collagen formation, individual fibres are cross linked by lysyl oxidase (through deamination of lysine and hydroxylysine residues) to form major

superstructures visible using light microscopy[260]. Further enzymes such as glycosyltransferases[261], hydroxylases[262], and isomerases[263] modify collagen fibrils to meet multiple tissue-specific needs by regulating density and rigidity of fibres. Glycine is found at almost every third residue and proline makes up approximately 17% of the collagen chain[264]. This amino acid in particular is important for enabling the packing of individual collagen molecules and they provide the stability required for formation of the fibre super helix.

Although the primary function of collagen is to provide structural support, it also plays fundamental roles in the regulation of tissue development, morphogenesis and homeostasis through its ability to mediate cell adhesion[265], proliferation[266], differentiation[267] and tissue regeneration[268]. In addition, it also regulates the density of the ECM as well as it's architectural composition, stabilising cell shape[269] and enabling communication between cells[270]. Although many of these functions are attributable to its mature, extracellular form, intracellular pro-collagen has also been reported to play a role in intracellular signalling[271]. Inherent mutations in specific collagen genes, affecting their transcription, post-translational modification, transport or folding generally result in severe pathologies and connective tissue disorders, including Bethlem myopathy, Ullrich congenital muscular dystrophy, Stickler syndrome, Marshall syndrome, Ehlers–Danlos syndrome, and others[272]. Many of these conditions cause pathological disturbance in multiple organ systems indicating the importance of specific collagens throughout the body. Natural degradation of collagen is directly linked to ageing[273] and therefore collagen is used for healthcare purposes such as cosmetics or burn medicine[274].

The role of collagens was described in colorectal and other cancers. In aggressive colon tumours, for example, collagen-rich stroma induced expression of mesenchymal genes and promoted tumour cell invasion[275]. In colon adenocarcinoma patients, high procollagen 11A1 expression was associated with nodal involvement, the development of distant metastases, and advanced Dukes stages[276]. Another study demonstrated that increased intratumoural expression of collagen XVII significantly correlated with higher TNM stage, infiltrative growth pattern, tumour budding, lymph node and distant metastasis in 141 patients with colon adenocarcinoma[277]. Interestingly, collagen chains IV and VI were decreased in colorectal cancer as compared to healthy control tissue[278]. Collagen type I promoted migration of colorectal cancer cells in a transwell assay through downregulating key genes of the Wnt/planar cell polarity (PCP) signaling pathway[279]. Further, formation of collagen I straps enhanced migratory potential of colon cancer cells SW480, SW620, and HCT116 cells, and X-radiation enhanced this effect[280]. Interestingly, inhibition of Wnt pathway was also described for collagen XVIII, which was found overexpressed in liver tumours[281]. In a 3D organotypic myoma invasion model, the loss of collagen XVII inhibited the migration and invasion of SCC-25 squamous cell carcinoma cells whereas there was no effect on the aggressive HSC-3 cells[282]. Overall, these findings implicate the ability of collagens to modify migratory properties of cancer cells and affect survival of patients with colorectal cancer.

3.1.2. Proteoglycans

Proteoglycans are a family of heavily glycosylated proteins consisting of a core protein with covalently attached chains of glycosaminoglycans; long, unbranched polysaccharides consisting of repeating disaccharide units[283,284]. As a major part of the ECM, proteoglycans are most commonly found in connective tissue, where they form

complexes with collagens[285] or hyaluronan[286]. The newly synthesised proteoglycan precursors enter the Golgi apparatus where they undergo glycosylation by glycosyl transferases, which recognise serine side chains of the core protein as a primer for growth of the polysaccharides[287,288]. These unbranched sugar chains make up to 60% of the proteoglycan`s total mass and define the functional specialisation of these molecules. Proteoglycans are involved in a wide variety of biological processes including cellular adhesion[289,290], structural maintenance[291,292] and signal transduction[293,294]. In particular, they play a role the actin cytoskeleton stability and focal adhesion under different conditions[295]. Proteoglycans are also able to sequester and subsequently release various cytokines and growth factors (for example FGFs, platelet-derived growth factor [PDGF], VEGF and connective tissue growth factor [CTGF]), thereby having a major role in regulation of communication between the ECM and cells (outside-in signalling)[295].

Multiple studies have indicated an aberrantly increased content of proteoglycans in malignant tumours, including colon cancer[296] and proteoglycans have been linked to cancer progression through several putative mechanisms. Increased levels of Versican were associated with higher microvessel density in metastatic colorectal cancer patients[297]. In addition, Versican expression correlated with both cancer relapse and poor patient outcomes in breast, prostate, and several other cancer types[298]. Besides Versican, a role in colon cancer has been indicated for other proteoglycans including Lumican, which expression was found associated with poor prognosis in patients with advanced colorectal adenocarcinoma with nodal metastases[299], and Biglycan, expression of which correlated with lymph node metastases, poor tumour differentiation and distant metastases[300]. In another study, both Lumican and Versican correlated with adenoma-to-carcinoma progression based on IHC analysis of human tissue specimens[301]. Further,

proteoglycan Decorin was demonstrated both *in vitro* and *in vivo* to inhibit pro-tumourigenic protein activin C, which enhanced colorectal cancer cell invasion proliferation and migration[302].

Little is known about the mechanisms through which these proteoglycans might contribute to cancer formation or progression; however, several studies indicate on the involvement of the integrin-TGF-β1 signaling pathway. In the melanoma microenvironment, Biglycan promoted invasiveness through the induction of integrin-β1 expression, leading to increased tissue stiffness[303]. In pancreatic cancer cells, TGF-β1 induced Biglycan expression through the Smad-activating function of ALK5 and GADD45bet[304]. Induction of Versican V2 expression using the antiarrhythmic drug amiodarone reduced the expression of EMT factors via AKT and ERK signalling, induced E-cadherin by inhibiting GSK3β, and also repressed the phosphorylation of EGFR at multiple tyrosine residues[305]. In addition, TGF-β1 promoted osteosarcoma cell migration and invasion via the miR-143-versican pathway *in vitro*[306]. Interestingly, loss of host stromal Versican induced angiogenesis in murine subcutaneous tumours, facilitated growth of xenografts, and activated TGF-β-mediated signal transduction[307]. Lumican produced by cancer associated fibroblasts increased gastric cancer cell proliferation, migration and invasion *in vitro* via integrin-β1-FAK signaling pathways[308]. Further, Nejad et al.[309] showed that breast cancer cell line 4T1 overexpressing Decorin exhibited a decreased TGF-β expression as well as elevated NF-κB DNA binding activity. Multiple proteoglycans, including Decorin, Biglycan and soluble Betaglycan, were reported to be cleaved by pro-apoptotic serine protease granzyme B to release active TGF-β1[310].

Thus, there is accumulating evidence that certain proteoglycans can have an effect on colon cancer cells` behaviour and worsen the prognosis in patients with colorectal adenocarcinoma.

3.1.3. Glycoproteins

Glycoproteins constitute a diverse group of glycosylated ECM proteins including Laminins, Fibronectins, Nidogen, Tenascins, Thrombospondins, and Elastin. They not only participate in ECM assembly but also act like reservoir of ECM-bound growth factors and ligands for cell surface receptors, mainly integrins[256]. Similar to other matrisome molecules, cleavage of glycoproteins can generate fragments with different functions than in their original full-length protein[256]. Glycoproteins are involved in a number pathological processes, including cancer. One of the most abundant glycoproteins, fibronectin, is upregulated in colorectal cancer stroma[311], where it activates VEGF-C secretion and promotes angiogenesis through the PI3K/Akt signaling pathway[312] and launches EMT of cancer cells by via the FAK-Src signaling pathway[313]. In addition, it was shown that increased matrix stiffness promotes the production of the extra domain-B splice variant of Fibronectin, which is strongly associated with highly angiogenic phenotype in tumours[314]. Some other splice variants of fibronectin can also affect cell adhesion and spreading[315]. Knockdown of Fibronectin 1 in two colorectal cancer cell lines, LOVO and SW1116, reduced cell proliferation, migration and invasion and activated cell apoptosis through downregulation of Bcl-2, MMP-9, Twist and upregulation of Bax, Caspase-3 and E-cadherin[316].

Another important member of the glycoprotein family is Tenascin-C, which also can be found in colorectal cancer stroma[317], where its expression predicts postoperative overall survival (P = 0.005; n=139 patients analysed), liver metastasis-free survival (P = 0.001; n=128 patients analysed)[318], and also correlates with EMT phenotype[319] and presence of circulating tumour cells[320]. Evidence from *in vitro* research suggests that Tenascin-C is produced by myofibroblasts and sends pro-invasive signals to human colon cancer cells through RhoA and Rac[321]. Regulation of Tenascin-C in human colorectal tumours is regulated by β-catenin/Wnt pathway[322], and in cultured myofibroblasts its expression is controlled by TGF-β[321]. In human brain tumour-initiating cells, Tenascin-C stimulated their proliferation via α2β1 integrin-mediated mechanism that upregulated Notch ligand Jagged1[323]. Further, glioblastoma cells treated by recombinant Tenascin-C produced pro-angiogenic factors promoting endothelial cell survival and tubulogenesis[324]. Supporting the role of Tenascin-C in metastatic seeding, overexpression of EMT regulators *Slug* and *Sox9* significantly upregulated the expression of Tenascin-C in MCF7 and MDA-MB-231 breast cancer cells, and knockdown of these genes reversed the observed upregulation[325].

Laminins are high-molecular weight (400 to 900 kDa) glycoproteins mostly involved in the formation of basal lamina. They are heterotrimeric molecules comprised from α-chain, β-chain, and γ-chain and found in five, four, and three genetic variants, respectively[326]. The expression of LAMC2 in colorectal cancer clinical samples correlated with poorer disease-specific, recurrence-free, disease-free, and overall survival in colorectal cancer[327], whilst LAMB1 was proposed as a promising serological marker for this disease[328]. Furthermore, high tissue expression of LAMB3 was a poor prognostic factor for stage III colorectal cancer patients (52.3 vs. 70.7% 5-year survival, P =

0.038)[329]. Another study showed that a decreased ratio of laminin-332 β3 to γ2 subunit mRNA is associated with poor prognosis in colon cancer[330]. Interestingly, inhibition of 67-kDa laminin receptor sensitised the colorectal cancer cell line SW620 to chemotherapy, suggesting potential clinical relevance of laminin blockade[331]. On the other hand, however, 67-kDa laminin receptor was reported to play a role in inducing cancer-selective apoptosis[332]. Collectively, these data indicate that glycoproteins are ECM components important for cancer progression and metastasis.

3.1.4. Aims

1. To generate a method to efficiently extract ECM proteins for analysis;
2. To compare the abundance of ECM proteins in liver metastasis and normal liver using proteomics;
3. To utilise bioinformatics tools for analysis of survival using proteomics data.

3.2. Results

3.2.1. Composition of the ECM from human colorectal liver metastases

3.2.1.1 ECM isolation

As previously noted, ECM research is hampered by its natural insolubility resulting from extensive cross-linking between the ECM protein constituents[88]. These characteristics greatly limit the set of tools and biochemical methods which can be used for extraction, solubilisation, and subsequent purification of ECM proteins. Work previously completed, such as my masters research project[250], focused on optimisation of methods and techniques for ECM extraction; we refined a method of a detergent-aided whole organ decellularisation in order to lyse and deplete the cellular components of a mouse liver with a view to developing a technique capable of decellularising liver metastases thereby allowing for analysis of the ECM at the metastatic niche. The use of detergents has a number of advantages, including high degree of decellularisation as well as being time efficient[255]. The latter is particularly important for preserving the presence of less stable proteinaceous elements of the ECM. Having tested multiple detergent-containing buffers, we identified that the mixture of 1% SDS and 0.01% NH_4OH prepared in double distilled water produced the most effective decellularisation of immersed normal mouse liver over a 72 h period (Figure 1).

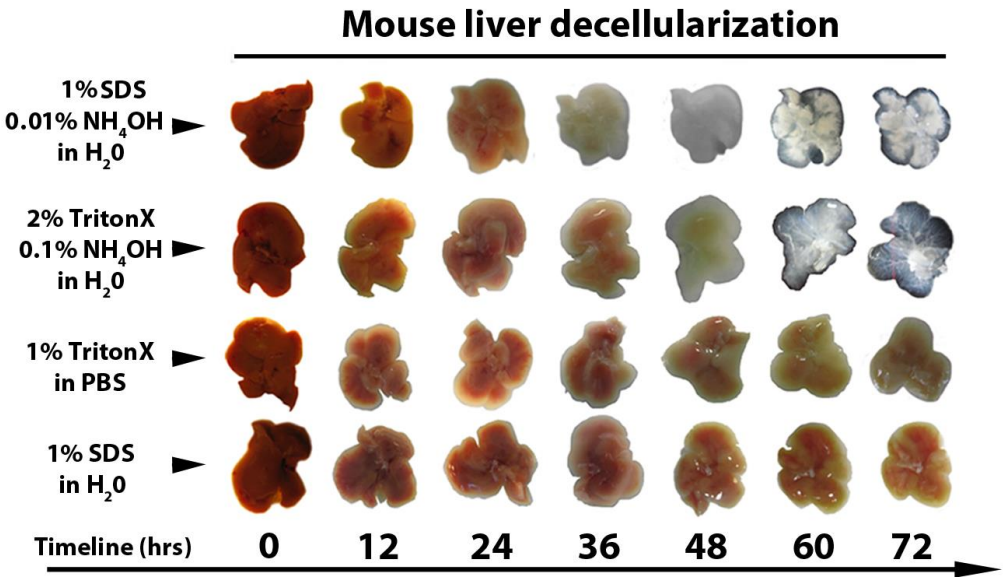


Figure 1 Tissue decellularisation. Representative images of mouse livers decellularised by the indicated solutions. Adapted from master's thesis.

Using this solution, we next turned our attention to the decellularisation of hepatic colorectal metastases removed from patients with curative intent. Adjacent uninvolved liver from matched patients was used as a control tissue (N=5 per group). The tissues were immersed in the decellularisation buffer for 72 h in order to obtain protein for proteomics analysis (see Figure 2 depicting experimental design).

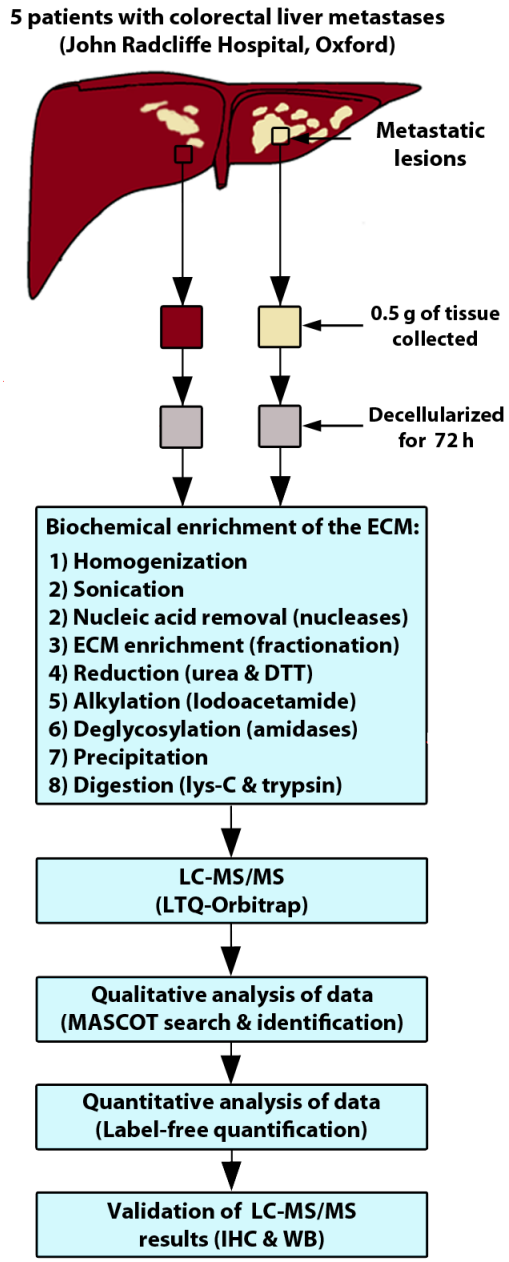


Figure 2 Experimental design.

To confirm the depletion of cellular components and retention of ECM molecules, we sectioned the resulting decellularised tissues and performed immunohistochemistry staining for H&E as well as immunofluorescent staining for key ECM proteins[333] such as fibronectin, collagen IV, collagen V, and pan-laminin, all counterstained with the DNA dye DAPI. As demonstrated in Figure 3, decellularised tissues were positive for all the ECM proteins yet devoid of nuclear staining, indicating successful depletion of cellular

components with retention of ECM proteins. Notably the decellularised tissues also retained a degree of their primary architecture.

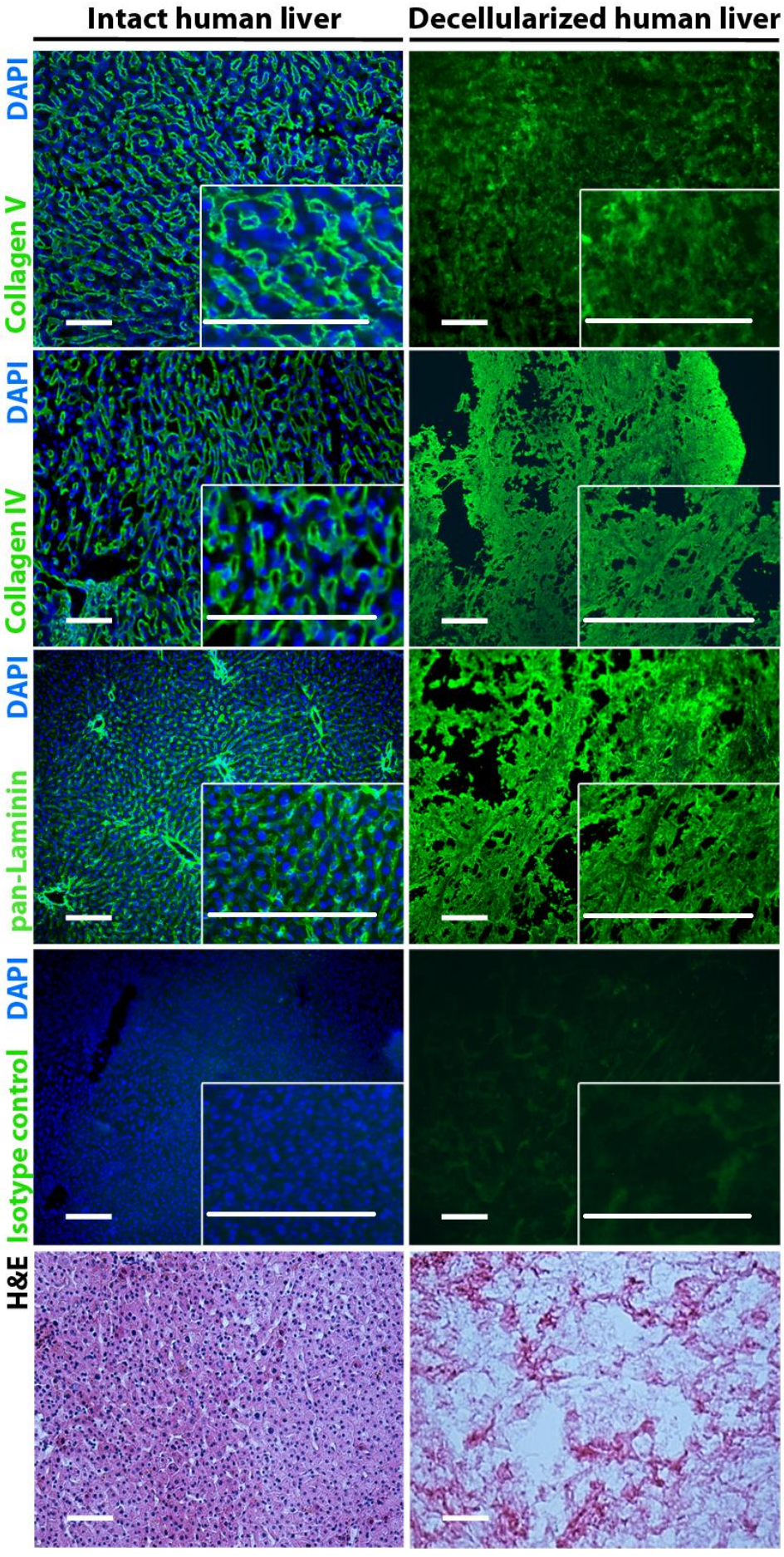

Intact human liver
Decellularized human liver
DAPI
Collagen V
DAPI
Collagen IV
DAPI
pan-Laminin
DAPI
Isotype control
H&E

Figure 3 Analysis of decellularised tissues. Human liver tissues, decellularised or intact, were stained for the indicated proteins, isotype control antibody, or H&E. Decellularised tissues were positive for all the ECM proteins yet devoid of nuclear staining, indicating successful depletion of cellular components with retention of ECM proteins. The decellularised tissues also retained a degree of their primary architecture. Scale bars = 100 µm. Experiment was performed once.

Consistently, the decellularised tissues were positively stained for eosin (stains extracellular protein), whereas no nucleus-specific haematoxylin stain was observed (Figure 3), again indicating that the given tissues were effectively decellularised by the detergent-containing buffer.

3.2.1.2. ECM enrichment

The step of tissue decellularisation was followed by further protein purification and removal of contaminating cellular components by a sequential process of subcellular fractionation, and enzymatic depletion of nucleic acids and oligosaccharides. This additional biochemical enrichment step was partially adapted from Naba et al.[251] and was required because we previously found that decellularised tissues contained high amounts of DNA, sugars, and low molecular weight intracellular proteins (<100 kDa), which suggested insufficient enrichment for ECM and made them incompatible with mass spectrometry analysis (data not shown). Finally, the samples were reduced, alkylated, precipitated using the methanol and chloroform method[334] and eventually digested with proteases (lys-C and trypsin) for the subsequent liquid chromatography and tandem mass spectrometry analysis. In order to confirm the successful enrichment of samples for ECM proteins, we performed SDS-PAGE separation followed by silver staining of protein bands (Figure 4). The increased proportion of high molecular weight proteins in the preparations confirmed enrichment of ECM proteins, as the ECM is composed of an abundance of high molecular weight proteins (>100 kDa), many of which are formed in the extracellular compartment after generation of smaller protein chains

within the cell. Thus, our biochemical enrichment methodology improved the purity of the ECM fraction, successfully eliminated the majority of intracellular proteins and thereby enabled the analysis of liver metastasis ECM by high-throughput proteomics.

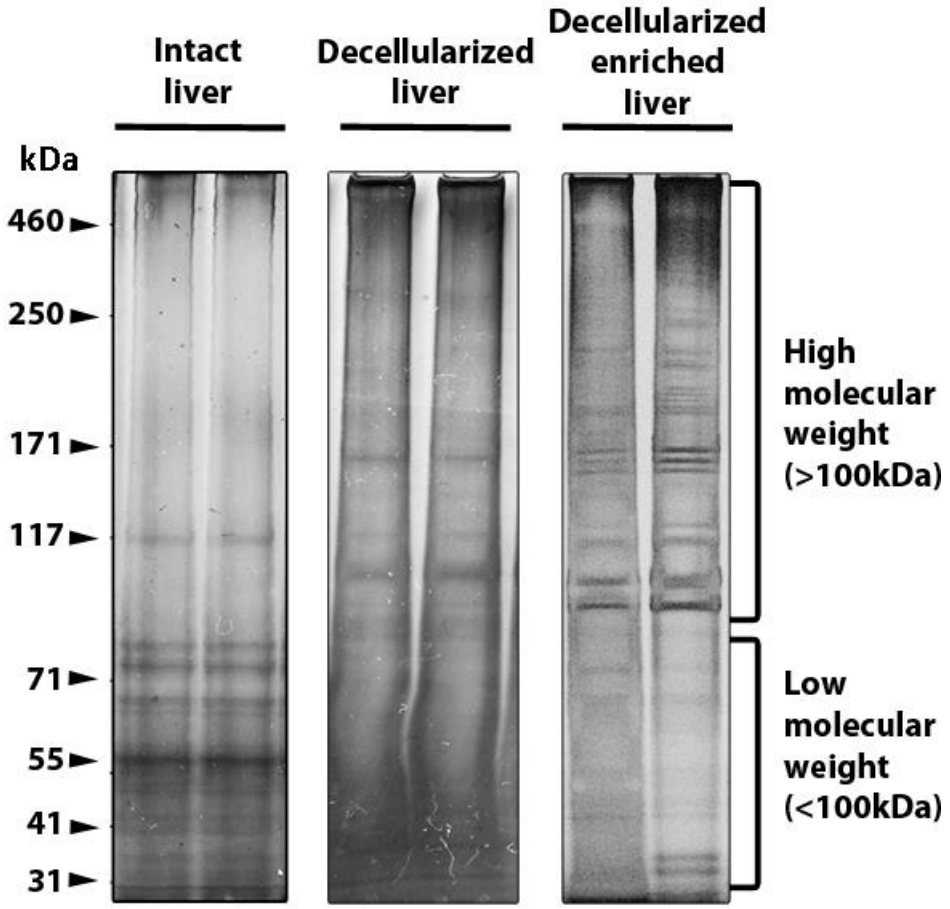


Figure 4 Analysis of ECM proteins by silver staining. Total protein extracted from intact (left), decellularised (middle), and decellularised plus biochemically enriched (right) human liver tissue adjacent to metastasis was resolved by SDS-PAGE and silver stained. High and low molecular weight areas were considered >100 kDa and <100 kDa, respectively. Decellularised and enriched liver has considerably more ECM proteins compared to other groups. Experiments were performed in parallel in duplicates. The experiment was performed once.

3.2.1.3. Proteomics

We next performed label-free mass spectrometry of decellularised liver metastases and normal livers in collaboration with the Target Discovery Institute (Dr.

Benedikt Kessler lab), University of Oxford. The initial qualitative analysis was performed using the emPAI scoring algorithm[335] and enabled to draw a preliminary picture of proteomics analysis. We identified a total of 1097 proteins, which were categorised in accordance with the dataset of complete human matrisome developed by Richard Hynes lab at Massachusetts Institute of Technology (web.mit.edu/hyneslab/matrisome/)[251]. As discussed above, this dataset was generated using the *in silico* approach to identify the definitive list of proteins belonging to the ECM based on the presence of conserved motifs characteristic of matrix proteins[251]. Out of 1097 proteins identified by proteomics, 158 were categorised as matrisome proteins, including 56 proteoglycans, 12 glycoproteins, 27 collagens, 17 ECM regulators, 30 ECM-affiliated proteins, and 16 secreted factors (Figure 5). For comparison, the human matrisome dataset developed by Naba et al.[251] includes 195 proteoglycans, 35 glycoproteins, 44 collagens, 171 ECM regulators, 238 ECM-affiliated proteins, and 344 secreted factors.

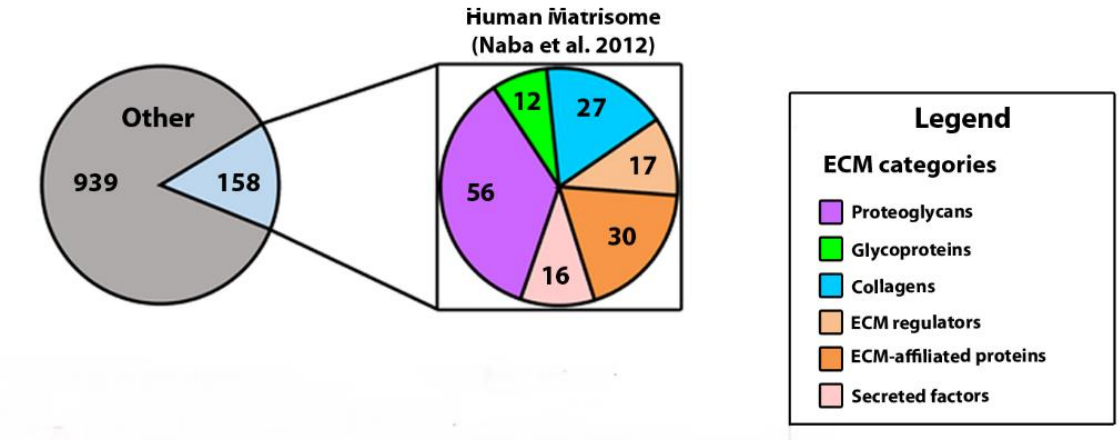


Figure 5 Qualitative proteomics results. Decellularised and ECM-enriched fractions obtained from human colorectal hepatic metastases and paired adjacent unaffected liver tissues (n=5 per group) analysed by LC-MS/MS. Left pie chart displays the total number of proteins identified by MASCOT search in at least 1 individual sample and by at least 2 peptides. Right pie chart displays the ECM categorisation of the identified proteins in accordance with Naba et al.[251]

Similar to Hynes, we distributed those into three groups, namely, (1) the core matrisome, encompassing ECM glycoproteins, proteoglycans, and collagens; (2) the ECM-associated proteins, comprising ECM regulators, ECM-affiliated proteins, and secreted factors; and (3) other proteins, which consists of non-ECM proteins from other fractions. Importantly, the proteomics results exhibited low inter-sample variability in protein content and considerable overlap of the matrisomal proteins, suggesting successful sample preparation (Figure 6A, B).

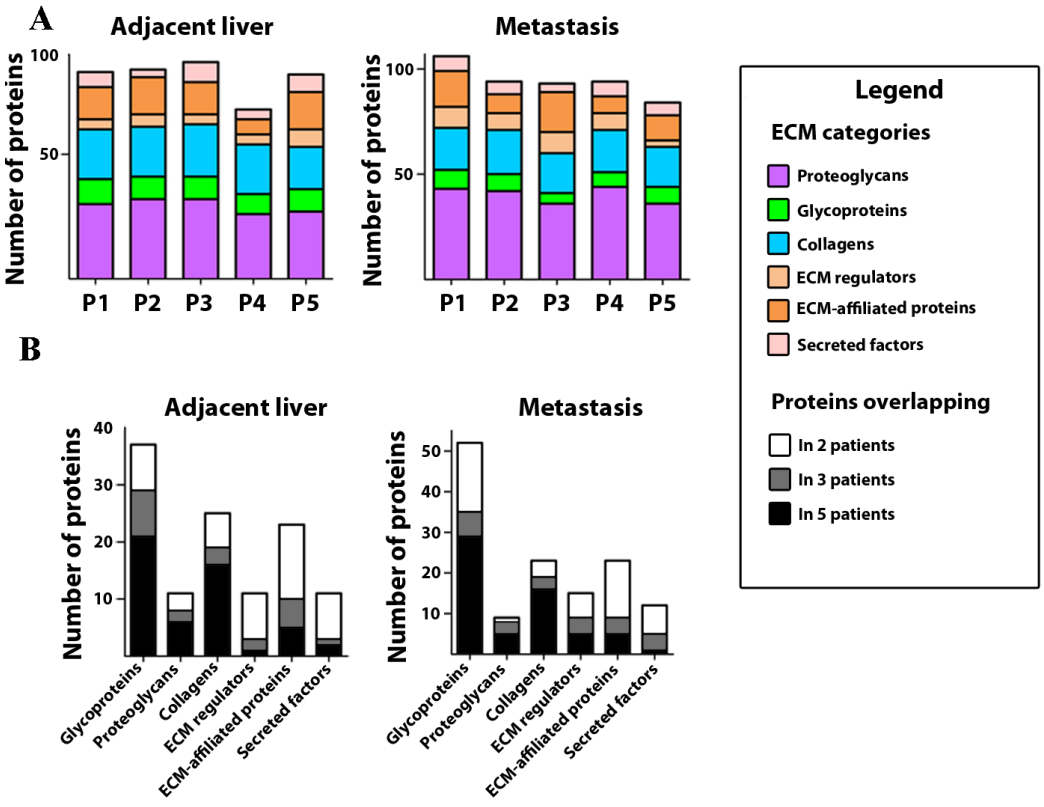
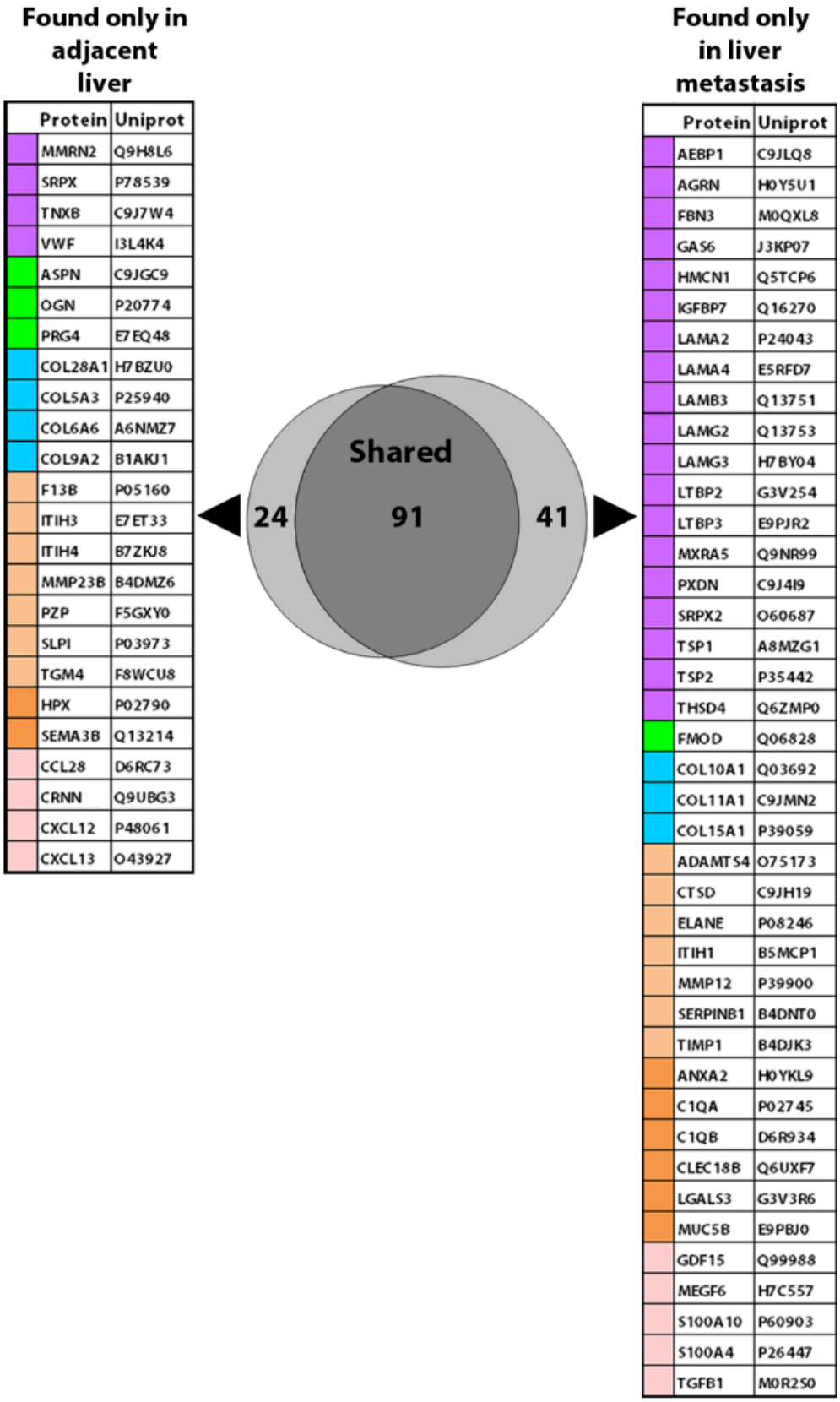


Figure 6 Inter-sample variability of proteomics results. A. Bar graphs show inter-sample variability between individual samples in metastasis and adjacent liver groups within each ECM category. B. Bar graphs show the overlap of identified proteins in patients across ECM categories. P – patient.

The majority of these matrisomal proteins were found in both uninvolved normal and liver metastasis tissues (91, at least 2 peptides); 24 proteins were exclusively present in the adjacent liver tissue, whilst 41 proteins were found only in liver metastases (Figure

7). The full list of proteins identified during the qualitative analysis can be found in the Supplementary Table 1.



Protein	Uniprot
MMRN2	Q9H8L6
SRPX	P78539
TNXB	C9J7W4
VWF	I3L4K4
ASPN	C9JGC9
OGN	P20774
PRG4	E7EQ48
COL28A1	H7BZU0
COL5A3	P25940
COL6A6	A6NMZ7
COL9A2	B1AKJ1
F13B	P05160
ITIH3	E7ET33
ITIH4	B7ZKJ8
MMP23B	B4DMZ6
PZP	F5GXY0
SLPI	P03973
TGM4	F8WCU8
HPX	P02790
SEMA3B	Q13214
CCL28	D6RC73
CRNN	Q9UBG3
CXCL12	P48061
CXCL13	O43927

Protein	Uniprot
AEBP1	C9JLQ8
AGRN	H0Y5U1
FBN3	M0QXL8
GAS6	J3KP07
HMCN1	Q5TCP6
IGFBP7	Q16270
LAMA2	P24043
LAMA4	E5RFD7
LAMB3	Q13751
LAMG2	Q13753
LAMG3	H7BY04
LTBP2	G3V254
LTBP3	E9PJR2
MXRA5	Q9NR99
PXDN	C9J4I9
SRPX2	O60687
TSP1	A8MZG1
TSP2	P35442
THSD4	Q6ZMP0
FMOD	Q06828
COL10A1	Q03692
COL11A1	C9JMN2
COL15A1	P39059
ADAMTS4	O75173
CTSD	C9JH19
ELANE	P08246
ITIH1	B5MCP1
MMP12	P39900
SERPINB1	B4DNT0
TIMP1	B4DJK3
ANXA2	H0YKL9
C1QA	P02745
C1QB	D6R934
CLEC18B	Q6UXF7
LGALS3	G3V3R6
MUC5B	E9PBJ0
GDF15	Q99988
MEGF6	H7C557
S100A10	P60903
S100A4	P26447
TGFB1	M0R2S0

Figure 7 Lists of ECM proteins identified exclusively in metastasis or adjacent liver (in at least 1 patient). There was 91 shared protein between groups

Interestingly, the majority (939) of the proteins identified by proteomics were not categorised as ECM-derived based on the matrisome dataset, which could suggest poor enrichment of the ECM; however, reclassification of these proteins with the Panther enrichment analysis system (http://www.pantherdb.org/)[252] identified the majority of them as ECM or exosomal components (Figure 8). This suggests that the presence of these proteins was not due to cellular contamination, but instead was the result of retention of exosomal contents by the ECM. Additionally, this could suggest that the matrisome dataset developed by Hynes is not fully inclusive and can be further expanded.

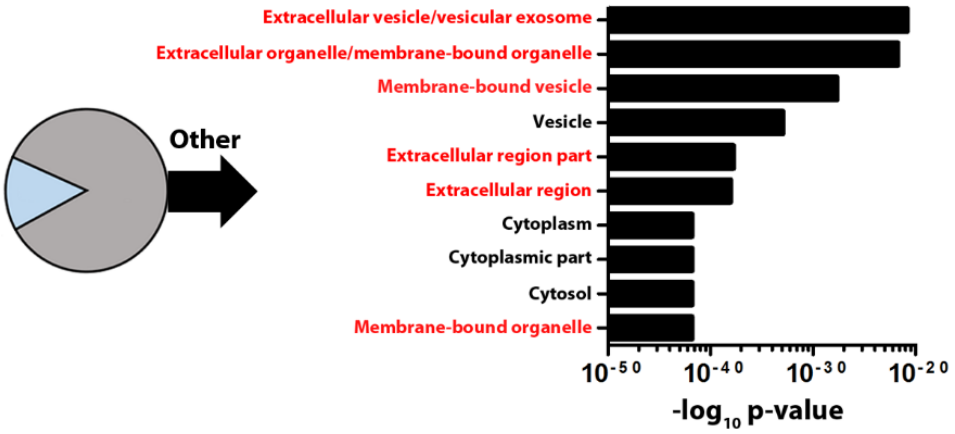


Figure 8 GO enrichment analysis. Proteins that were not categorised as ECM proteins were processed through Panther GO enrichment analysis. Top 10 enriched GO categories are shown. ECM-related GO categories are highlighted in red.

We then applied quantitative label-free analysis to the proteomic data using Progenesis QI software in order to quantitatively analyse the composition of metastatic ECM. This software has been successfully used for the accurate measurement of compounds in complex mixtures, including protein solutions[336]. Principal component analysis revealed distinctive differences between the ECM from metastatic colorectal cancer and adjacent liver (Figure 9A). Clustering analysis identified 287 proteins with

statistically significant differential abundance between the normal liver and metastases (two-way ANOVA, $P < 0.05$) (Figure 9B).

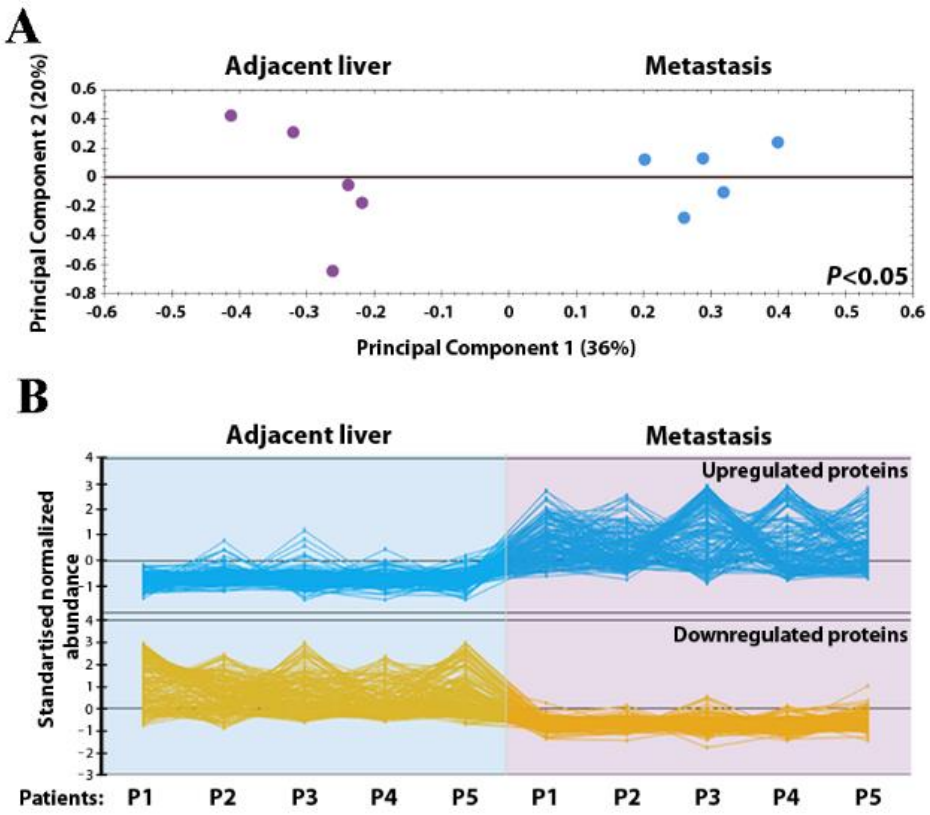


Figure 9 Principal component and clustering analyses of proteomics results. A. Decellularised and ECM-enriched fraction from colorectal hepatic metastases and paired adjacent unaffected liver tissues (n=5 per group) was analysed using LC-MS/MS with label free quantitation. Shown is the principal component analysis graph of the relative protein abundance between metastasis and adjacent liver groups ($P < 0.05$). Percentage of variance is shown in brackets. B. Clustering analysis graph shows proteins differentially abundant in metastasis and adjacent liver groups.

We next obtained a list of 70 proteins with fold change of >3 that were significantly ($P < 0.05$, FDR applied) up- or downregulated in the ECM from liver metastases compared to that from adjacent liver (Figure 10). Many of the upregulated proteins have previously been linked to metastatic progression, including Versican, TIMP1, LTBP1-3, DDR1 and S100A10[337–341]. Interestingly, multiple proteins from this list were found to be members of TGF-β pathway, which is particularly relevant in colon

cancer[86] and fibrotic/ECM response[342]. The full list of proteins detected and quantified by the Progenesis QI software can be found in the Supplementary Table 2.

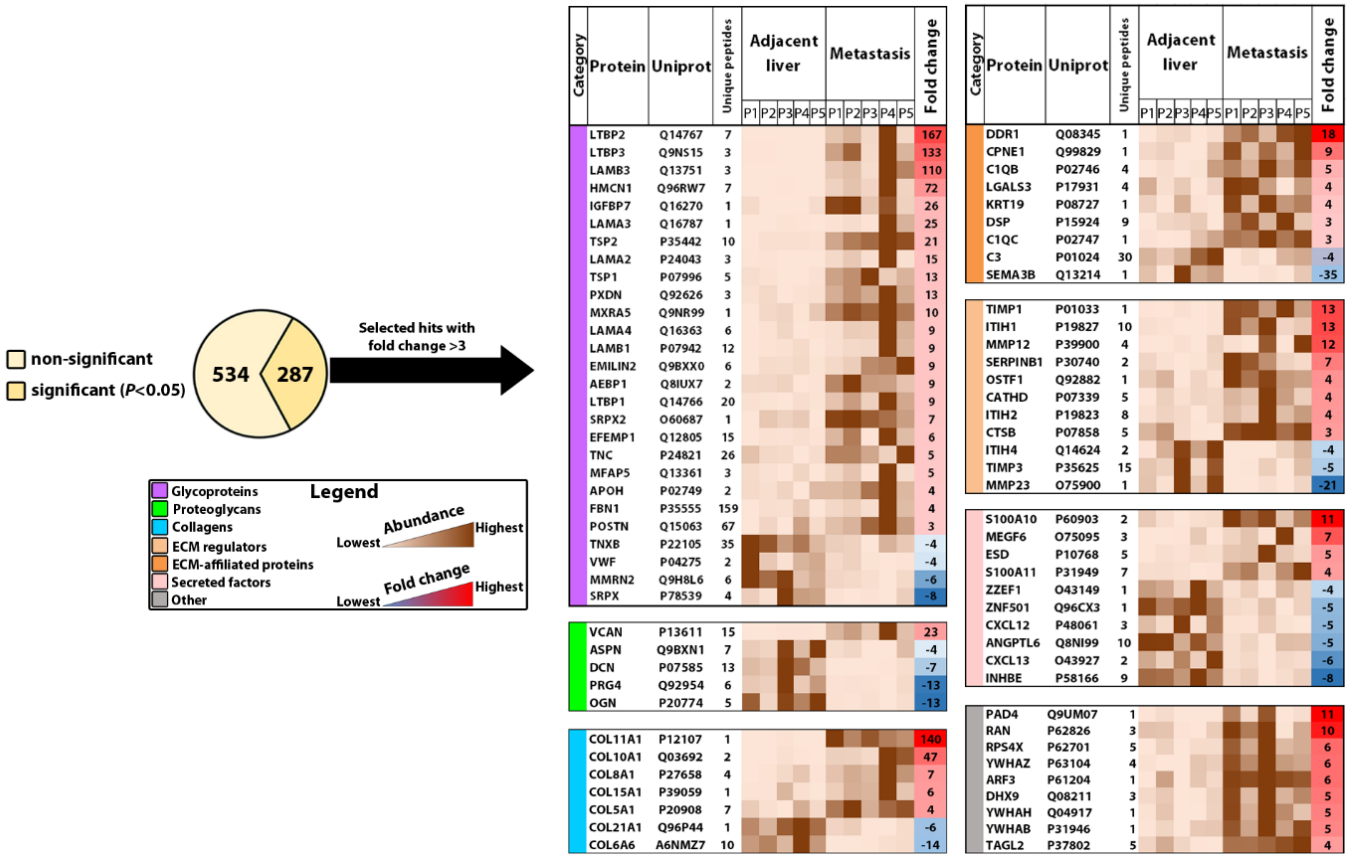


Figure 10 Quantitative proteomics results. Statistically significant protein hits (287 out of 821 hits, two-way ANOVA, P < 0.05) were classified either as "the ECM" (69 hits) or "other" proteins (218 hits). A heat map was constructed based on hits which were restricted to a fold change threshold of more than 3 and ranked in accordance with ECM category and fold change.

3.2.1.4. Validation of proteomics

To confirm the proteomics findings, the expression of six proteins were assessed using immunoblotting and immunostaining on an independent set of resected hepatic metastases from colorectal cancer patients (Figure 11A-D). Differential expression similar to the proteomic results was confirmed for all the proteins tested; Versican, Tenascin, S100A11 and COL5A1 were overexpressed, whilst Matrix Metalloproteinase-23 and Mimecan were downregulated in the metastases compared to adjacent liver. This

successfully validates the results generated by mass spectrometry in an independent cohort of patients.

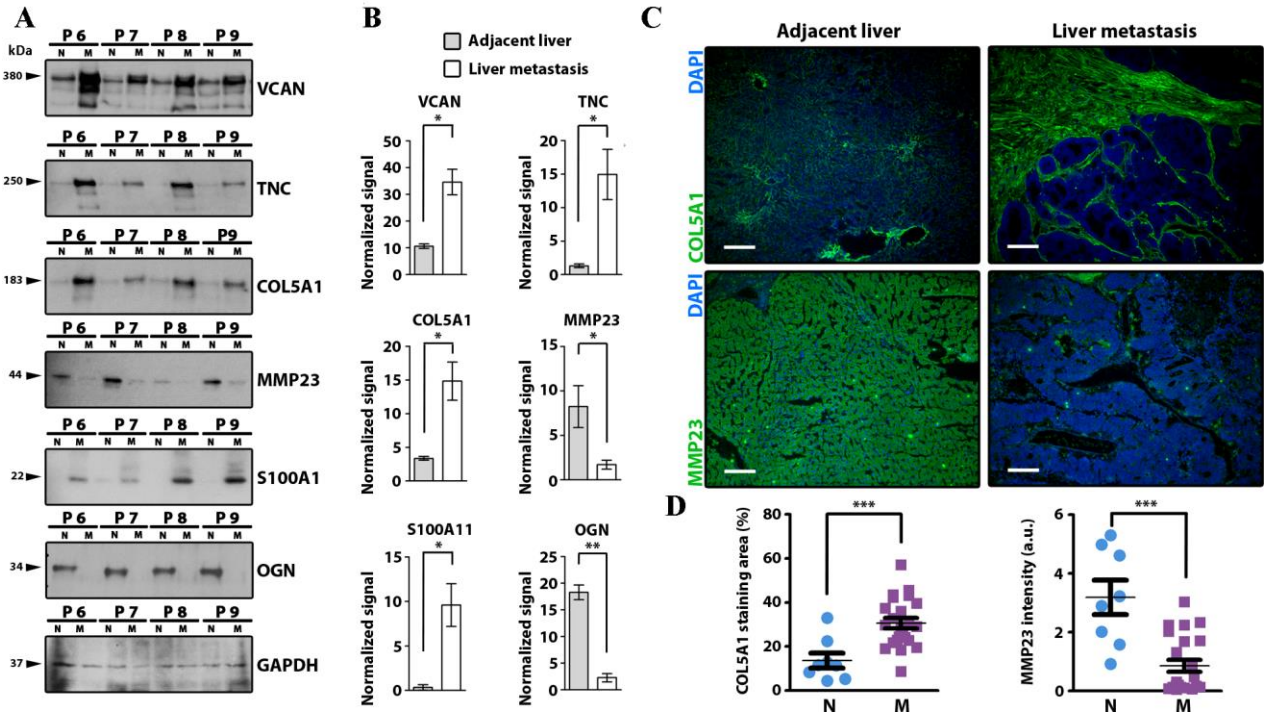


Figure 11 Validation of proteomics. A. Tissue lysates extracted from human colorectal hepatic metastases and paired adjacent unaffected liver tissues (*n*=4 per group) were immunoblotted for the indicated proteins. Adjacent liver and metastasis tissues are marked N and M, respectively. GAPDH was used as a loading control. B. Densitometry analysis of protein bands detected in A. Densitometry was normalized to loading control. C. Representative images of COL5A1 and MMP-23 staining of 23 colorectal liver metastasis tissues (right) and 8 non-matched adjacent unaffected liver tissues (left). 10x objective was used. Scale bar = 100 µm. D. Quantification of COL5A1 staining area and MMP-23 intensity from the experiment C. At least 2 fields of view were analysed per patient. For B, error bars indicate s.e.m. (P* < 0.05, ***P* < 0.01, two-tailed paired Student's t-test). For D, error bars indicate s.e.m. (****P* < 0.001, two-tailed unpaired Student's t-test). Experiments were performed once.**

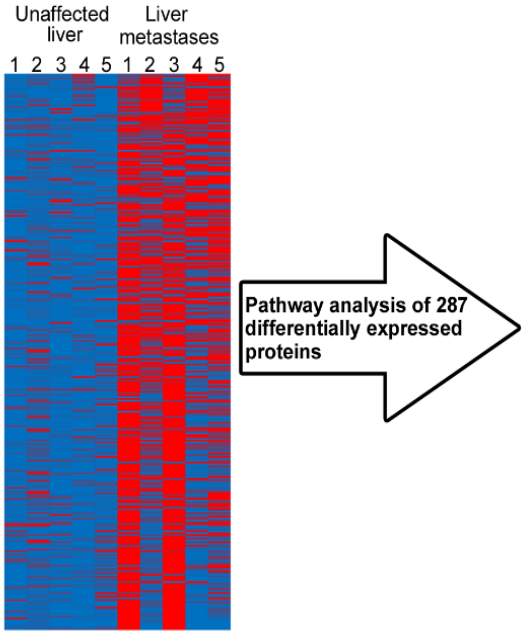
To conclude, the aforementioned experiments demonstrate that decellularisation and biochemical enrichment are successfully used as a method for ECM preparation for proteomics analysis. Through this process we have identified a distinctive ECM composition in colorectal cancer liver metastases relative to normal liver. Finally, we

corroborated the proteomics results using immunoblotting and immunostaining on a different set of resected liver metastases and uninvolved tissues.

3.2.2. Identification of the ECM gene signature predicting survival from colon cancer

3.2.2.1. Bioinformatics analysis of ECM proteomics results

The stroma plays a role in colon cancer progression as evidenced by the finding that stroma-rich tumours are associated with reduced disease-specific survival[343]. Although we have identified a set of ECM proteins over-represented in liver metastases comparative to normal liver, it is unclear whether these proteins play a role in colorectal cancer progression. We next wished to determine whether this was the case and so developed a bioinformatics pipeline to identify whether ECM molecules from our



Pathway	Source	Overlapping genes/total pathway genes	Enrichment P-value (after correction)	Disease-free survival HR (95%CI) [p-value]
ECM organisation	Reactome	74/261	1.06×10^{-25}	**1.55 (1.02-2.36) [0.04]**
β1 integrin cell surface interactions	PID	34/66	2.02×10^{-20}	**2.22 (1.42-3.47) [0.0005]**
ECM receptor interaction	KEGG	31/87	2.34×10^{-13}	**2.17 (1.38-3.42) [0.0008]**
Integrins	INOH	37/124	2.36×10^{-13}	**1.60 (1.02-2.52) [0.04]**
Focal adhesion	KEGG	48/207	6.50×10^{-13}	1.52 (0.89-2.6) [0.13]
Chaperone-mediated protein folding	Reactome	23/49	8.62×10^{-13}	1.01 (0.52-1.95) [0.98]
Platelet degranulation	Reactome	29/81	1.24×10^{-12}	1.52 (0.95-2.41) [0.08]
Collagen biosynthesis and modifying enzymes	Reactome	23/65	6.59×10^{-10}	1.21 (0.73-2.00) [0.46]
Elastic fibre formation	Reactome	17/35	9.54×10^{-10}	1.44 (0.86 to 2.40) [0.17]
Laminin interactions	Reactome	12/23	2.60×10^{-07}	**1.74 (1.06-2.87) [0.03]**

Figure 12 Differentially expressed proteins identified by mass-spectrometry (left) were processed by the pathway enrichment analysis. Top 10 enriched pathways (right) were analysed using the cBioportal software[253] to identify whether overexpression of these pathways affects colon cancer survival (TCGA mRNA data). Red values indicate statistically significant differences in disease-free survival hazard ratio (HR) for pathways overexpressed in colorectal cancer.

proteomics data are associated with reduced colorectal cancer survival. To do so, we analysed the dataset generated by quantitative proteomics using the pathway enrichment analysis. Initial pathway analysis of the 287 differentially expressed ECM proteins identified enrichment of several molecular pathways from this protein set (Figure 12). Expectedly, the top three enriched pathways were closely related to the ECM component: β1 integrin cell-surface interactions, ECM receptor interaction, and ECM organisation (Figure 12).

Using the publicly available cancer survival analysis software cBioportal[253] (http://www.cbioportal.org/), we then compared the disease-free survival of colorectal cancer patients who overexpressed these molecular pathways and those that did not (Figure 12).

Strikingly, overexpression (z-score +1.5) of each of the pathways was associated with significantly reduced disease-free survival, indicating the enrichment of ECM gene identified in liver metastasis results in poor outcome from primary colorectal cancers (See disease free HR in Figure 12). We next identified all the genes which were common between these top three pathways (β1 integrin cell-surface interactions, ECM receptor interaction, and ECM organisation) and thus developed a 13 gene signature of commonly overrepresented ECM genes identified in liver metastasis and associated with adverse survival (Figure 13).

Figure 13 13 gene signature. 13 genes commonly found in the top three enriched pathways (β1 integrin cell-surface interactions, ECM receptor interaction, and ECM organisation). These are all the genes that were common to the top three enriched pathways.

As expected, overexpression of these 13 genes in combination predicted reduced disease-free survival in three large, independent colon cancer datasets including the TCGA dataset (Figure 14). Thus, here we identified the 13 gene ECM signature predicting survival from colon cancer, highlighting new avenues for prognostication of this disease.

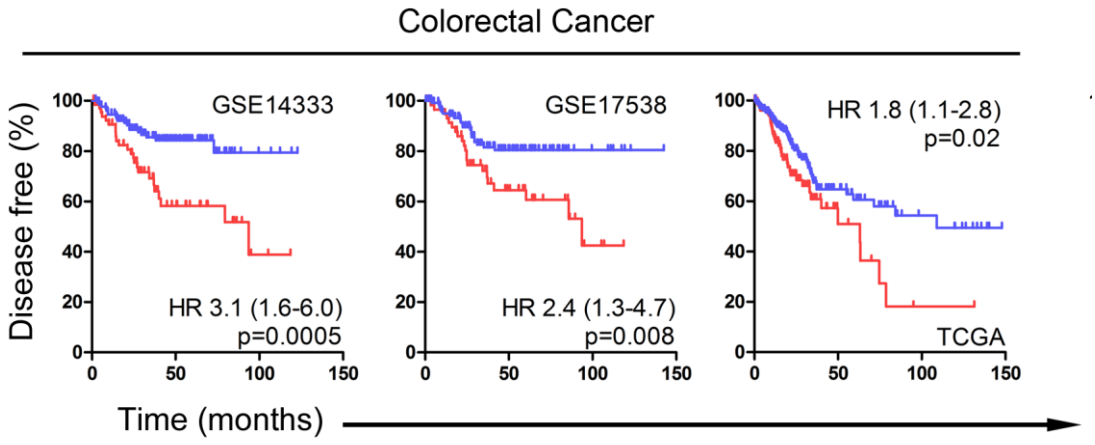


Figure 14 13 gene signature predicting survival from colon cancer. Disease-free survival from colon cancer in patients overexpressing the 13 gene signature based on three different mRNA datasets (GSE14333, 290 patients, GSE17538, 177 patients, and TCGA, 640 patients).

3.2.2.2. *Identifying the cellular source of 13 gene signature*

We next turned our attention to the potential cellular source for the 13 ECM genes in the signature. To this end, experimental liver metastases were generated by intrasplenic injection of GFP^+ MC38 murine colon cancer cells. We chose this cell line because it reproducibly generates liver metastases in a short-term period of 12-14 days[344] and resulting tumours were comprehensively characterised earlier by other lab members[345,346]. Mouse livers were perfused with collagenase *in situ* and metastases were resected following density gradient separation of hepatic stellate cells from the liver[347], preparation of single cell suspension, and FACS-sorting of four cell populations: cancer cells (GFP^+ $CD45^-$ $CD31^-$), immune cells ($CD45^+$ $CD31^-$ $CD140\alpha^-$), endothelial cells ($CD31^+$ $CD45^-$ $CD140\alpha^-$), and fibroblasts ($CD31^-$, $CD45^-$, $CD140\alpha^+$) (Figure 15).

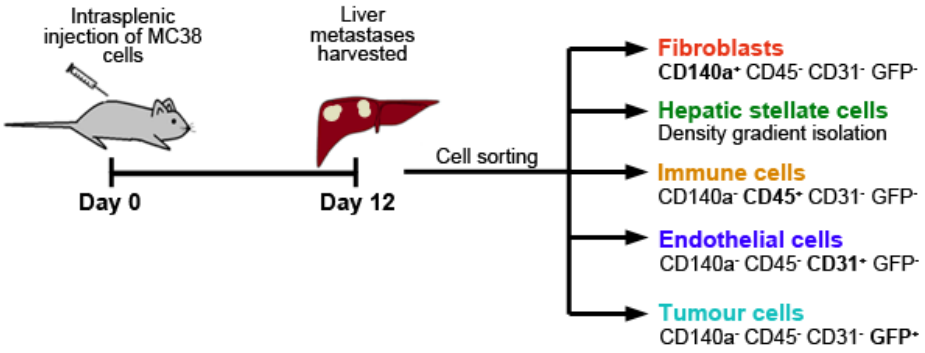


Figure 15 Experimental setup for isolation of different cells from mouse liver metastatic microenvironment.

The isolated cells were lysed and mRNA extracted. To confirm the specificity of sorting, we first profiled the resulting cells for characteristic genes specific for these cell populations (Figure 16A). For example, isolated endothelial cells expressed Pecam1

(CD31) and von Willebrand factor and but showed no expression of Itgam (CD11b) or PTPRC (CD45), whilst sorted immune cells demonstrated the opposite expression pattern.

Having profiled the aforementioned cell populations for 13 gene signature, we identified that fibroblasts and hepatic stellate cells where the predominant source of these genes (Figure 16B). This result was somewhat expected, as cancer-associated fibroblasts have long been known to be the major producer of tumour stroma[86]. Hepatic stellate cells are able to differentiate into fibroblasts upon external stimulus[348] and have similar functions to hepatic fibroblasts[349], so their contribution to the production of the ECM signature is also not surprising.

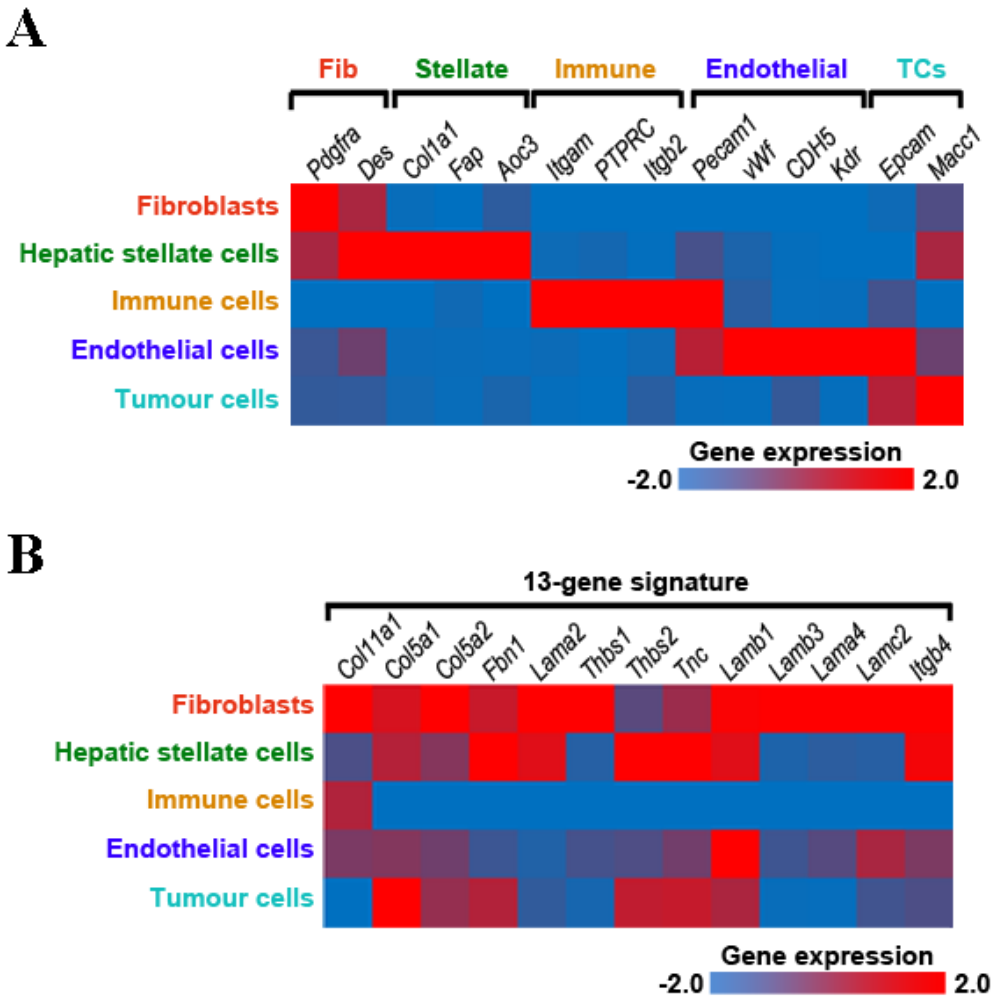


Figure 16 mRNA profiling of sorted cells. A, B. mRNA from the indicated cell lines was probed for the indicated genes using the quantitative PCR. Resulting delta CT values were transformed into relative expression values and plotted as a heat map.

To further confirm that the 13 gene signature derives from activated cancer-associated fibroblasts, we co-stained one of the proteins comprising the signature, Thbs1, with a well-recognised marker of fibroblasts ER-TR7 and a marker of cancer-associated fibroblasts alpha smooth muscle actin (αSMA) in a set of resected liver metastases specimens. The expression of Thbs1 colocalised with both markers (P = 0.0043), supporting our PCR results demonstrating expression of this ECM gene in fibroblasts from the liver metastasis microenvironment (Figure 17A, B).

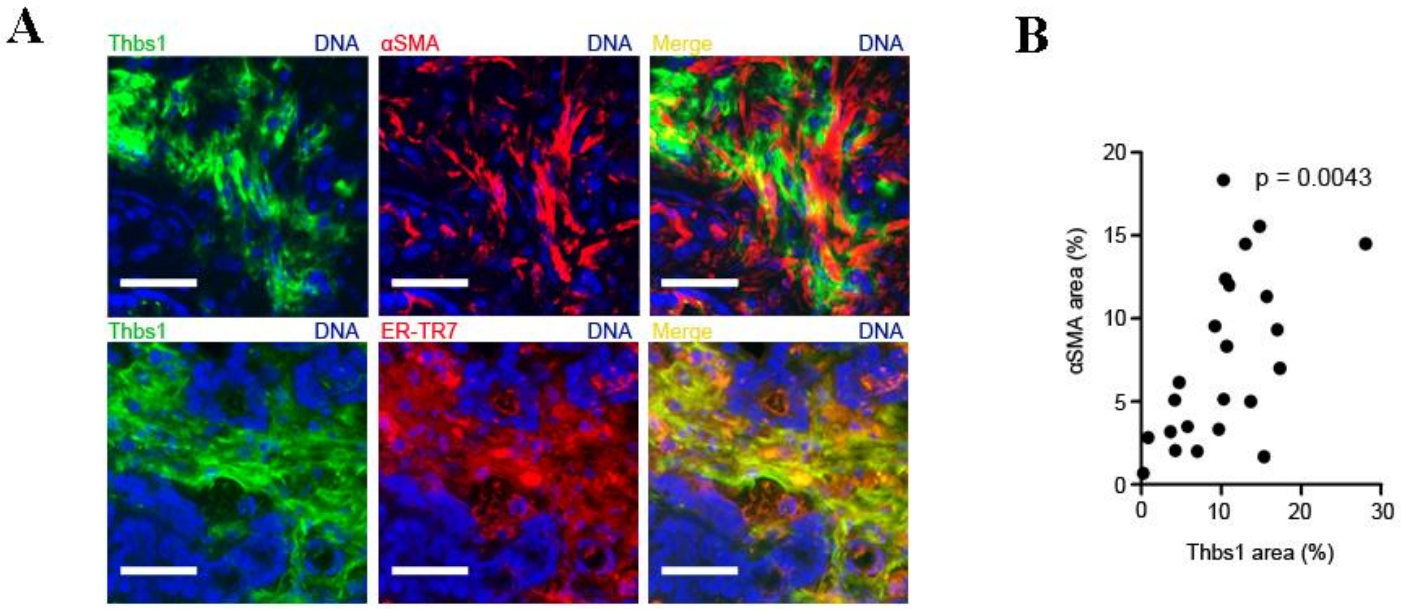


Figure 17 Staining of human liver metastases for Thbs1, ER-TR7 and αSMA. A. A set of 23 resected human liver metastases was immunostained for the indicated proteins. All sections were counterstained with DAPI. Scale bar = 100 µm. B. Pearson`s correlation analysis of the staining area of the indicated proteins.

Little is known about the ability of cancer cells to drive the production of ECM from fibroblasts. To demonstrate that cancer cells are able to promote the production of ECM genes in fibroblasts, we treated cultured mouse embryonic fibroblasts (MEF) with conditioned media collected from MC38 cells. The cells were lysed at different time points and mRNA extracted. The majority of ECM genes identified in our 13 gene signature demonstrated a sharp rise in mRNA expression levels for cells treated with MC38-conditioned compared to control media between 2 and 4 h post treatment (Figure 18A), indicating that MC38 cancer cells are able to promote the production of ECM genes in fibroblasts. Similar results were obtained upon the addition of MC38 conditioned media into cultured hepatic stellate cells isolated from mouse (Figure 18B, C). Thus, MC38 colon cancer cells at least in part drive the production of the 13 genes in MEF fibroblasts and isolated hepatic stellate cells.

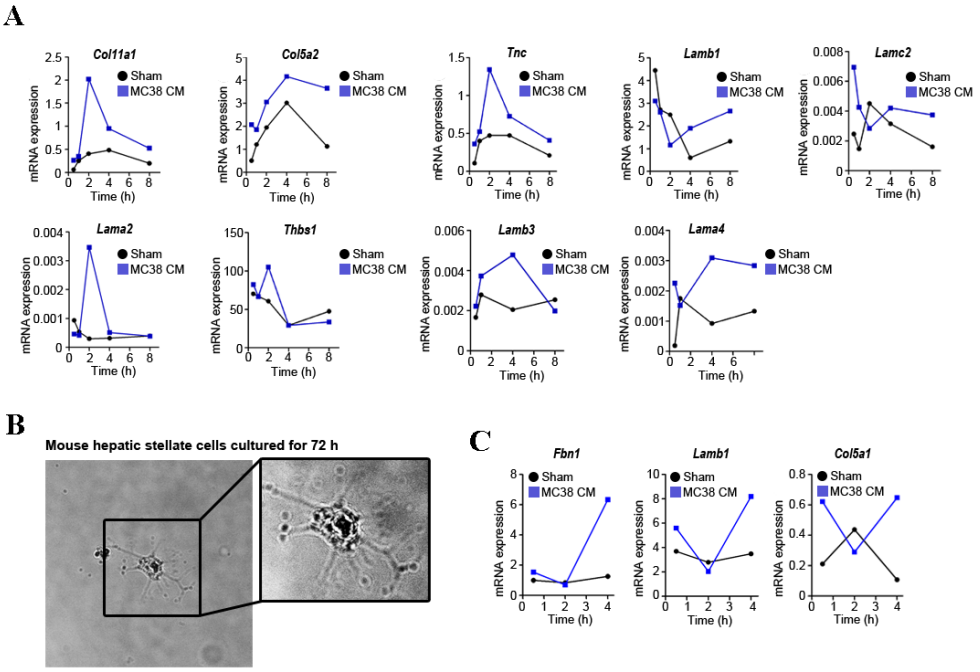


Figure 18 Profiling of ECM genes in MEFs or hepatic stellate cells. A, C. Cultured MEFs or freshly isolated mouse hepatic stellate cells were serum starved for 4 h, followed by the addition of MC38 conditioned media or control serum-free media. The cells were collected at the indicated time points and profiled for the indicated genes using the quantitative PCR (normalised to HPRT). B. Bright field image of an isolated mouse hepatic stellate cell after 72 h of culture. Other genes constituting the 13 genes signature did not show a difference in expression relative to sham.

3.2.2.3. Exploring the signalling pathways driving the 13 gene signature expression

We then examined which molecular pathways are activated in fibroblasts in response to MC38 conditioned media. Having profiled an array of phosphorylated kinases using the dot blot method, we identified that signal transducer and activator of transcription 3 (STAT3) and mammalian target of rapamycin (mTOR) were the most prominently phosphorylated factors in MEFs (at sites Y705 and S2448, respectively) incubated in MC38 conditioned media (Figure 19).

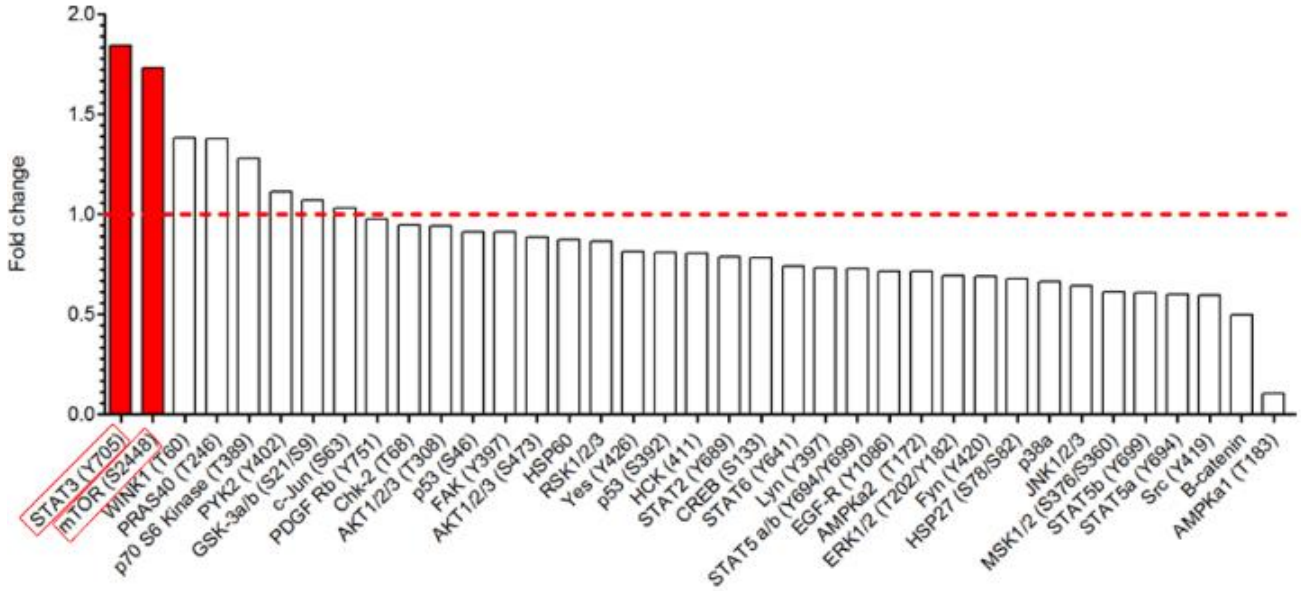


Figure 19 Pathway analysis in MEFs treated with MC38 conditioned media. Cultured MEFs were serum starved for 4 h, followed by the addition of MC38 conditioned media or control serum-free media. The cells were incubated for 2 h and protein extracted. The resulting protein lysates were profiled for phosphorylated and non-phosphorylated proteins using the commercially available Proteome Profiler Human Phospho-Kinase Array Kit (some of the proteins were non-phosphorylated only). Graphs indicate the fold change of the ratio of phosphorylated to non-phosphorylated proteins. STAT3 and mTOR are highlighted in red.

Both STAT3 and mTOR are known to be deregulated in multiple cancers[350–352]; in the presence of MC38 conditioned media as compared to vehicle control confirming that mTOR signalling promotes the expression of ECM genes in response to MC38-conditioned media (Figure 20C).

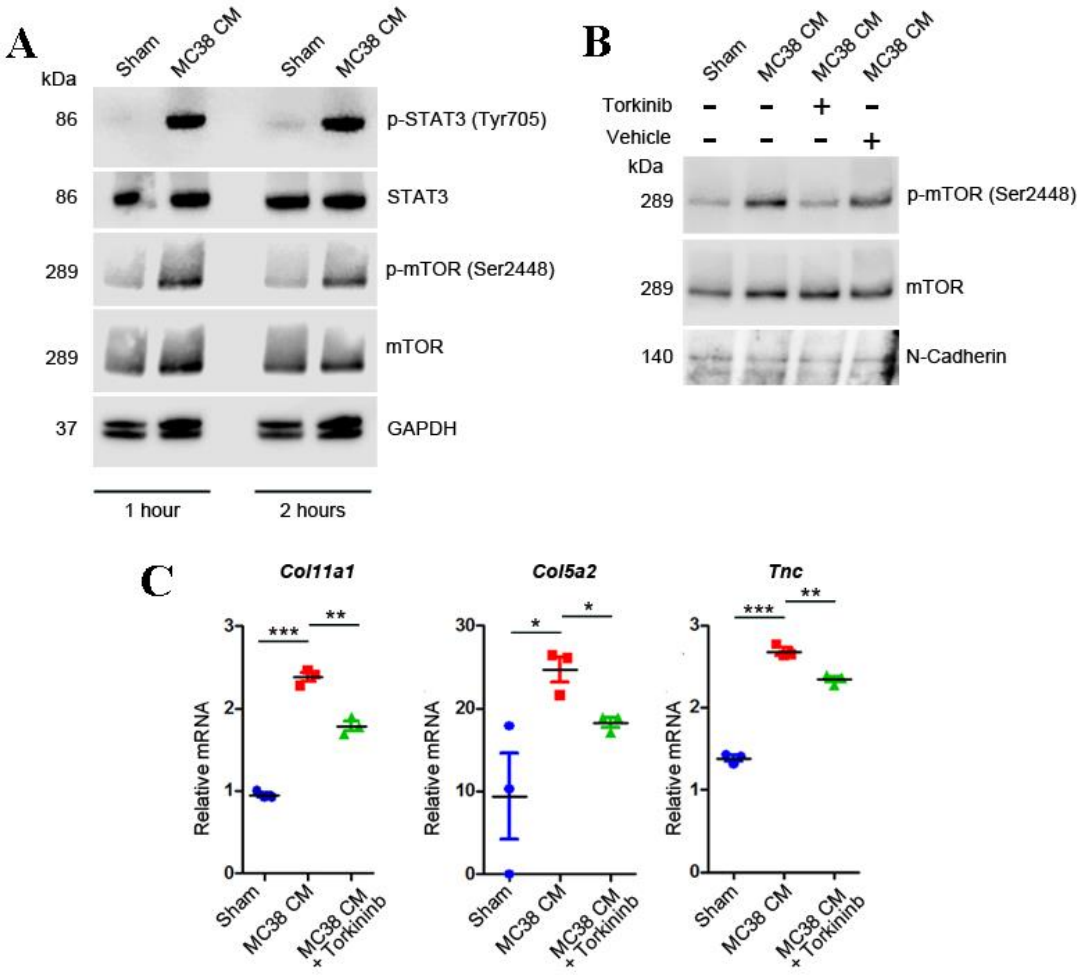


Figure 20 mTOR inhibition in MEFs. A, B. Cultured MEFs were serum starved for 4 h, followed by the addition of MC38 conditioned media or control serum-free media (or conditioned media with 2.5 μM mTOR inhibitor torkinib or vehicle in B). The cells were incubated for 1 or 2 h and protein and mRNA were extracted. The resulting protein lysates were profiled for the indicated proteins. C. mRNA expression levels (normalised to HPRT) of the indicated genes in MEFs from experiment in A (2 h timepoint).

In conclusion, through bioinformatics analysis of proteomic data derived from decellularised human liver metastases, we identified 13 ECM genes, overexpression of which is associated with significantly worsened disease-free survival in colorectal cancer patients. Through analysis of murine liver metastases, we demonstrate that fibroblasts

and hepatic stellate cells are the predominant source of the 13 gene signature in the tumour microenvironment. In addition, we show that conditioned media from MC38 cells at least partially induces the expression of the 13 gene signature in cultured MEFs and isolated mouse hepatic stellate cells. Finally, MC38 conditioned media activates the phosphorylation of STAT3 and mTOR in fibroblasts and pharmacological inhibition of mTOR phosphorylation reduces the conditioned media-driven expression of ECM genes.

3.3. Discussion

Despite the fact that excessive ECM deposition and tumour-associated desmoplastic stroma are recognised hallmarks of cancer, the specific set of proteins comprising the tumour ECM and their potential impact on cancer biology and prognosis are less well studied. Extensive protein crosslinking and insolubility make separation of the ECM from cellular contents complicated and mass spectrometry-aided analysis of the tumour ECM is therefore extremely challenging. In this chapter for the first time we show a novel approach for high-throughput quantitative proteomics of tumour ECM, in which efficacy was independently validated by classical methods of molecular biology. To isolate the ECM, we have chosen the method of detergent-aided whole organ decellularisation which retains the intact ultrastructure of the matrix architecture as well as the macromolecular composition of the ECM. There are many other methods for decellularisation, including biological (enzymatic and non-enzymatic) and physical (thermal and electrical) approaches. Biological methods are generally not sufficient for the decellularisation of a whole organ, whilst physical approaches are damaging to the tissue and can negatively influence the integrity and morphology of the ECM due to heat. Equally, these methods result in an excessive amount of cellular debris retained within the tissue.

To choose the appropriate decellularisation buffer we have previously performed several optimisation experiments, which demonstrated that 0.1% SDS and 0.01% NH_4OH in water is the optimal formulation to use. We have also optimised the protocol for biochemical enrichment of decellularised matrices, which resulted in a high yield of the ECM content, as demonstrated by immunohistochemistry results as well as silver staining of SDS-PAGE-resolved ECM proteins. The results from this chapter demonstrate that decellularisation followed by biochemical enrichment yielded superior purification of tissue ECM allowing more quantitative analysis. These methods should be widely applicable to future studies of cancer ECM.

Despite the fact that the variability in core matrisome content was small between samples, the groups of secreted factors and ECM-affiliated proteins exhibited considerable variation in the number of proteins identified in each sample. These results are consistent with our previous findings published in my masters thesis[250] as well as the results by Naba et al.[251] where inter-sample variation for these groups was reported to be approximately 50%. This can be explained by the fact that many of the non-structural ECM proteins (i.e. non-core matrisome proteins) are especially prone to wash out during sample preparation and loosely bound to the fibres of the matrix. Furthermore, these protein categories are generally present at a low molar concentration within the ECM, and therefore maybe not be easily detectable by mass spectrometry, which could also be a reason for variation in these groups.

It is important to note that only a small proportion of proteins (15%) identified by mass spectrometry were categorised as matrisomal. On the one hand, this could be

because of poor purification and insufficient extraction of the ECM component. On the other hand, it can be explained by limitations of the complete matrisome dataset, which is only based on the *in silico* prediction rather than careful functional evaluation of proteins with regards to their actual relation to the ECM. Analysis of protein bands presented in Figure 4 suggests that decellularisation followed by biochemical enrichment is very effective in purification of high molecular weight proteins (>100 kDa), therefore it is unlikely that over 85% of protein content in these samples were non-ECM molecules. In order to test whether the complete matrisome dataset is reliable in terms of protein classification, we have performed an enrichment analysis of proteins not categorised as ECM, with regards to their cellular localisation based on GO data. As demonstrated, the Panther classification system identified those proteins as highly enriched for ECM-related GO annotations, such as extracellular vesicles (top 1), extracellular organelle (top 2), and extracellular region (top 5). Unexpectedly, these results also indicated a potential major influence of exosomal components on the ECM microenvironment. Exosomes have previously been found to play a role in progression of multiple primary cancers as well as metastatic disease[353], but the crosstalk between cancer-associated exosomes and the ECM is not reported. Thus, our results suggest that the complete matrisome dataset is not fully inclusive and may require further extensions, raising questions about the definition of the term "ECM". The isolation of the ECM in the study by Naba et al.[251] was performed by cellular fractionation without decellularisation or other enrichment methods, which could explain potential drawbacks of their method leading to less accurate results.

Application of the quantitative label-free algorithm enabled us to identify differences in abundance of ECM proteins in the normal uninvolved liver and colorectal

cancer liver metastases. Principal component analysis demonstrated a clear separation between study groups, indicating that liver metastases have a distinctive ECM composition (Figure 9). We identified groups of proteins either up or down regulated in the metastatic tissue as compared to the uninvolved liver. The final list of statistically significantly differentially expressed proteins included 70 molecules, the majority of which were overexpressed in the tumour tissue. It is interesting that many of these proteins demonstrated a high variation between metastasis samples, yet still being significantly different compared to the uninvolved liver. For example, patient 4 (metastasis) has a remarkably higher concentration of almost all ECM molecules as compared to other patients. This can be explained by inaccuracies during the quantitative analysis or different molecular background of colon cancer in these patients. As discussed above, there are believed to be four molecular subtypes of colon cancer: inflammatory, canonical, metabolic and mesenchymal, of which the latter is characterised by an extensive ECM deposition[85]. It is possible to speculate that the patient 4 has a mesenchymal molecular subtype of colon cancer which explains such extensive ECM deposition in his/her metastases.

Notably, the proteomics results indicated a significant contribution of TGF-β signalling pathway to the liver metastases ECM, as multiple members of TGF-β pathway were found remarkably upregulated. TGF-β signalling has been recognised as critical to colorectal cancer liver metastasis[354] and we found upregulation of the TGF-β binding proteins LTBP1-3, TSP1 and TSP2, all of which contribute to the activation of TGF-β[355,356]. TGF-β has recently been associated with cancer suppression through epithelial-mesenchymal transition, and its post-translational modifications were reported to alter its activity in the extracellular microenvironment[357,358]. The complex balance between these

signals and the function of epithelial-mesenchymal plasticity is still not completely understood.

Our proteomics results showed that many of the proteoglycans (as well as some of the collagens) were significantly down regulated in the tumour tissue compared with the uninvolved liver. This could be explained by liver-specific expression of these proteins, which is supported by multiple literature sources identifying proteoglycans as important players in liver physiology[359,360]. For example, heparan sulphate proteoglycan sequesters major modulators of lipoprotein metabolism such as lipases and apolipoproteins in the liver, enabling hepatic lipoprotein uptake and regulation of plasma lipid level[360].

High throughput analyses of biomedical data frequently produce inaccurate or even erroneous results due to complexity of the methodology or inappropriate statistical analysis[361]. It is therefore should be noted that in this chapter we performed an independent validation of our quantitative proteomics results. On a different set of resected liver metastases specimens, we confirmed that Versican, Tenascin, S100A11 and COL5A1 were overexpressed, whereas Matrix Metalloproteinase-23 and Mimecan were down-regulated in the metastases compared to uninvolved liver. We validated this using two classical methods of molecular biology, immunoblotting and immunohistochemistry, which indicates the reliability and trustworthiness of our proteomics results.

Cancer progression is directly associated with reduced disease-free survival[362] and deaths in colorectal cancer patients are primarily the result of metastatic disease. To

identify the set of ECM proteins responsible for progression of liver metastases, we utilised a bioinformatics approach applied to publicly available cancer survival data. First, we performed an enrichment analysis of molecular pathways on the entire proteomics dataset generated from ECM specimens. This enabled us to identify ECM-related molecular pathways which correlated with poorer survival from colon cancer. Having selected a common set of genes belonging to these pathways, we established a 13 gene signature which predicted survival from colon cancer in three independent datasets with a high level of statistical significance. Therefore, we speculate that these proteins are associated with cancer progression and are required in some way for the development of liver metastasis. We suggest that their combination could represent a helpful biomarker for tumour prognostication. However, it should be noted that the results presented in this chapter are only correlative and from its analysis we are unable to provide a mechanistic link between the expression of the ECM 13 gene signature and specific biological processes. It would be of interest to further investigate which aspects of cancer progression are influenced by the 13 gene signature. In addition, future research should address several questions unanswered in this chapter, namely, how could these proteins be helpful in the metastatic niche, and is there evidence for their function at the metastatic site?

We were also interested in identification of the cellular source of the 13 gene signature in the hepatic metastatic microenvironment. Using a mouse model of experimental metastases, we isolated cell populations that constitute this microenvironment, namely, cancer cells, immune cells, endothelial cells, fibroblasts and hepatic stellate cells. Importantly, we validated the quality of cell sorting by performing the quantitative PCR profiling genes specific for the above-mentioned cell populations.

Expectedly, we identified fibroblasts as the major source of these ECM genes; whilst hepatic stellate cells also partially contributed to the 13 gene signature. Cancer-associated fibroblasts have long been implicated in enhancing tumour growth, angiogenesis, inflammation and metastasis formation[363]. Unlike normal fibroblasts, they are actively recruited by tumour cells to generate desmoplastic stroma[364], in which they exhibit higher proliferation and migration rate[365]. It has also been reported that tumour cells can 'educate' cancer-associated fibroblast in a secondary site to form a niche for metastasis formation[364]. By altering the ECM composition, cancer-associated fibroblasts contribute to the malignant phenotype by altering cell-matrix interactions controlling the sensitivity of epithelial cells to apoptosis; producing and secreting proteins which in turn control epithelial cell survival and proliferation; influencing cell-cell interactions whereby cells are able to exchange membrane fragments and communicate; creating physical barriers due to generation of desmoplastic stroma, which facilitate the acquisition of drug resistance; and finally, activating epigenetic plasticity in neighboring cells[364,366,367].

Because cancer-associated fibroblasts are activated by the tumour, we hypothesised that the 13 gene signature expression in fibroblasts is driven by factors secreted by cancer cells into the tumour microenvironment. *In vitro* we showed that many of the genes comprising the 13 gene signature are indeed induced in MEFs 1 to 2 h after treatment with media conditioned by mouse MC38 colon cancer cells. Similar results were shown for isolated and cultured mouse hepatic stellate cells. Thus, we demonstrated that the 13 gene signature is produced predominantly by fibroblasts educated by cancer cell conditioned media. It remains to be determined, however, which factor is directly responsible for that effect; cancer cells are known to produce a plethora of cytokines, chemokines and growth factors[345], and identification of a molecule driving the production

of the 13 gene signature may open new perspectives in therapeutic targeting of desmoplastic stroma-producing cells in a subset of patients with mesenchymal subtype of colon cancer.

To provide an insight into the molecular pathways activated in MEFs in response to cancer cell conditioned media, we examined the phosphorylation status of several signalling molecules activated in MEFs the same experimental *in vitro* setup. Our findings implicate the involvement of STAT3 and mTOR phosphorylation as a result of treatment with conditioned media. Both of these molecules have been reported to be deregulated in various cancers, including colorectal adenocarcinoma[350–352,368–370].

Our work also identified mTOR and potentially STAT3 activity as an important driver of ECM gene expression in fibroblasts. These observations are supported by the research literature. Indeed, a recent study by Chakraborty[371] established STAT3 as an important molecular checkpoint for tissue fibrosis. The authors observed accumulation of phosphorylated STAT3 in fibrotic skin, and interestingly, its activation was mediated by TGF-β. Another study showed that STAT3 and interleukin 6 (IL-6) are activated in hepatic stellate cells by hepatic leukaemia factor (HLF) to promote liver fibrosis in human patients[372]. In a model of obstructive nephropathy following ureteric ligation, pharmacological inhibition of STAT3 reduced fibroblast activation and attenuated tissue fibrosis[373]. Furthermore, a large amount of evidence implicates STAT3 as a central mediator of idiopathic pulmonary fibrosis as reviewed by Prele et al.[374] With regards to mTOR, in human lung fibroblasts the mTOR signalling pathway has shown to be crucial for TGF-β mediated deposition of collagen[375]. In a different study, interstitial myofibroblasts from fibrotic kidneys exhibited increased phosphorylation of mTORC1,

whilst its pharmacological and genetic inhibition inactivated these cells[376]. Suppression of the mTOR pathway has also been suggested for the treatment of cardiac fibrosis[377] and pulmonary fibrosis[378] in the clinical setting. In our study, tumour-derived factors induced phosphorylation of Ser2448 residue of mTOR, which is known to be phosphorylated via the PI3 kinase/Akt signaling pathway[379], suggesting its upstream activation in this setting. Collectively, these and our findings are suggestive of the involvement of both mTOR and STAT3 in desmoplastic stroma formation and more specifically in the stimulation of the 13 gene ECM signature expression.

In summary, this chapter provided a proteomics-based characterisation of human liver metastases ECM and demonstrated a set of 13 ECM genes significantly predicting survival in patients with colon cancer. The following chapter is concerned with the mass spectrometry-aided analysis of post translational modifications, specifically citrullination, of human liver metastases ECM, and describes a potential mechanism through which citrullination may promote hepatic metastasis growth.

CHAPTER 4. CITRULLINATION FACILITATES COLORECTAL LIVER METASTASES BY ENFORCING MESENCHYMAL-TO-EPITHELIAL TRANSITION

4.1. Introduction

4.1.1. Mass spectrometry as a high-throughput method in biomedicine

The emergence of high-throughput technologies, such as multiplex ELISA, next-generation sequencing, proteomic or PCR arrays, and high-resolution mass spectrometry provoked a significant advance in biomedical research, enabling identification and investigation of various signaling pathways involved in both physiological and pathological processes. Development of these techniques allowed determination of molecular mechanisms of multiple human diseases, including endocrine disorders[380,381], autoimmune disorders[382,383], chronic obstructive pulmonary disease[384], atherosclerosis[385,386], chronic kidney disease[387,388], neurodegenerative disorders[389,390] and cancer[391–393]. In addition, findings generated by high-throughput technologies are used for drug screening purposes as well as network biology[394–398]. Implementation and subsequent application of such high-throughput methods often requires a collaboration of professionals working in different science areas, including omics technologies, computational biology, bioinformatics, life and health sciences, biotechnology and bioengineering.

In the field of cancer research, mass spectrometry has greatly contributed to the understanding of molecular basis of this disease in the post-genomic era. Currently, there are two main avenues in cancer proteomics research: (a) global-expression proteomics,

which simultaneously identifies a large number of proteins differentially abundant between cells, tissues, or body fluids in a quantitative manner, and (b) functional and cell mapping proteomics, which identifies interactions between proteins, protein and DNA, protein and RNA, or posttranslational modifications[399]. In this thesis we have explored both of these avenues. Whilst chapter 3 addressed the composition of human liver metastatic ECM, this chapter will be devoted to identification, quantification and investigation of a posttranslational modification, namely citrullination, in the same setting.

4.1.2. Epithelial-mesenchymal plasticity in the context of cancer metastasis

The transition of cells from epithelial to mesenchymal phenotype and vice versa is a well-documented phenomenon, playing a role in regulating migratory and invasive behaviour of cells under homeostatic and pathological conditions[400]. This transition is characterised by transformation of cells with epithelial phenotype, i.e. well-developed apico-basal polarity, defined cell-cell adhesions and established cell-matrix interactions, into the mesenchymal phenotype with high invasiveness, absence of junctions and front-back polarity. Importantly, the process of EMT is flexible and highly volatile, being able to rapidly change towards one or another end within the same group of cells in a short time period. Multiple studies describe the state of partial EMT, whereby cells existed in a hybrid state with an intermediate phenotype[401,402]. The mechanism of EMT is required for embryonic development, morphogenesis, fibrosis and wound healing; however, it has also been reported to contribute to such pathological processes as cancer progression, metastatic dissemination and seeding[403]. Implementation of the EMT program is governed by multiple transcription factors, to name a few, Snail1/2,

Zeb1, Twist etc.[404], as well as post-transcriptional regulators such as micro RNAs (miR) 200 and 34[405,406]. Notably, all aspects of transcriptional, post-transcriptional, and epigenetic control of EMT regulation are still not completely understood[407]. The most commonly used markers for EMT detection in cells and tissues are E-cadherin, Keratins, Occludins and Claudins for the epithelial phenotype and N-cadherin, Vimentin and Zeb1 for the mesenchymal phenotype. Here we will briefly review the roles of EMT in homeostatic and pathological conditions.

4.1.2.1. EMT in development and response to injury

Early genetic studies of *Drosophila* and vertebrate embryos revealed the crucial role of EMT in development. Mouse embryos lacking the *Snail* gene do not survive gastrulation[408], and the formation of parietal endoderm from primitive endoderm and visceral endoderm is primarily driven by the *Snail* gene[409]. During the process of implantation, trophoblasts localised at the tip of chorionic villi require EMT to invade the endometrium[410]. Primary mesenchyme formation in sea urchin embryo is directly controlled by Snail- and Twist-mediated EMT program execution[411]. In the same species, Snail inhibits the transcription of the *E-cadherin* gene and activates cadherin endocytosis along with promoting the delamination of primary mesenchyme cells by EMT[412]. These are only a few examples of many demonstrating a role of EMT in early development.

Another major function of EMT is regulation of repair-associated events, such as wound healing, organ fibrosis and tissue regeneration[413,414], where mesenchymal cells are required to produce large amounts of ECM. In mouse models of pressure overload and chronic allograft rejection, TGF-β was shown to promote mesenchymal phenotype in

epithelial cells, whilst bone marrow protein 7 (BMP-7) reversed this transition[414]. In a mouse model of kidney inflammatory injury, macrophages and activated resident fibroblasts underwent EMT through release of TGF-β, PDGF, epidermal growth factor (EGF), and FGF2[415] in order to induce repair processes. In a different study, mouse renal fibrosis was attenuated by administration of recombinant BMP-7, which improved renal function, histology, and animal survival[416]. The authors showed that these effects were mediated by BMP-7-driven reversal of the TGF-β-induced loss of the key epithelial player E-cadherin[416]. Another study demonstrated that elevated tissue-type plasminogen activator (tPA) levels were crucial in renal interstitial fibrogenesis through a cascade of events leading to MMP-9 induction, basement membrane destruction, and eventually, activation of EMT[417].

4.1.2.2. EMT in cancer

Currently there is no commonly accepted and established theory explaining how cancer cells launch the EMT program. It is hypothesised that at a primary tumour site, a subset of cancer progenitor cells (or cancer stem cells) undergo EMT and metastasise following clonal expansion. However, it is not known whether all progenitor cells undergo EMT simultaneously or different clones within the tumour progress towards EMT at different times[403]. Indeed, EMT activators (e.g. Twist1, Zeb1, Snail, etc.) have been reported to induce stemness properties and vice versa[418–420]. In either event, primary tumour cells require EMT or partial EMT in order to gain invasive properties and successfully metastasise. This is supported by multiple studies reporting high levels of EMT inducing factors in metastasizing cancer cells[421]. However, some of the EMT-inducing molecules, such as TGF-β, can also possess an anticancer effect by inducing apoptosis in tumour cells[422], although certain EMT activators were reported to inhibit

the programmed cell death[423]. EMT in cancer is also characterised by formation of invadopodia - specialised membrane protrusions mediating ECM degradation through release of MMPs. In particular, Twist1 has been reported as a key mediator of invadopodia formation in response to Snail-driven EMT-inducing signals[424]. Another evidence comes from imaging studies investigating the extravasation dynamics of metastatic cancer cells, which demonstrated that Twist1 expression in breast tumour cells induced cancer cell extravasation through forming large dynamic membrane protrusions in a zebrafish model[425]. Interestingly, activation of Twist1 was shown to be driven by EGF and STAT3[424], connecting EMT with the results presented above in chapter 3. A large body of evidence summarizing the cross-talk between STAT3 and activation of EMT program was recently reviewed by Wendt et al[426]. In addition, expression of EMT factors has been reported to activate stemness by promoting Ras, Wnt and Notch signalling pathways[427,428]. In particular, Zeb1 activity suppressed stemness-inhibiting microRNAs (miRNAs) such as miR-203 thereby promoting more mobile and migrating cancer stem cells[427].

Recent studies suggest the importance of mesenchymal-to-epithelial transition (MET), the reversal of EMT, in cancer metastasis formation[429]. It was shown that the EMT-MET switch is required for cancer cells to secure their place in metastatic sites and give rise to colonies, as demonstrated by Ocana et al.[430] who showed that simultaneous knockdown of EMT factors Twist1 and Prrx1 increases number of lung metastases after tail vein injection of BT-549 breast cancer cells. Another piece of evidence comes from the study by Tsai et al.[431], who demonstrated that activation of Twist1 in murine primary skin tumours increases the number of circulating tumour cells but not the number of lung metastases as compared to mice with doxycycline-mediated knockdown of this factor. It

was further shown that Twist1 inhibition is important for colonizing cancer cells to overcome EMT-associated growth arrest[431]. Thus, EMT and MET play a central role in progression and metastatic dissemination of cancers.

3.1.3. Aims

1. To investigate whether liver metastasis show increased levels of protein citrullination compared to unaffected tissue;
2. If yes, provide an insight on the role of citrullination in liver metastasis.

4.2. Results

4.2.1. Citrullination of human hepatic metastases

4.2.1.1. ECM proteins are predominantly citrullinated in liver metastases

Besides quantitative analysis of protein content, modern mass spectrometry enables the detection of post-translational modifications, including citrullination. We compared the amount of protein citrullination in the ECM from human liver metastases and adjacent normal liver using our proteomics results. To do so, we utilised a PEAKS® software pipeline[432] to search for citrullinated arginine residues in the samples previously analysed by LC-MS/MS (Figure 21).

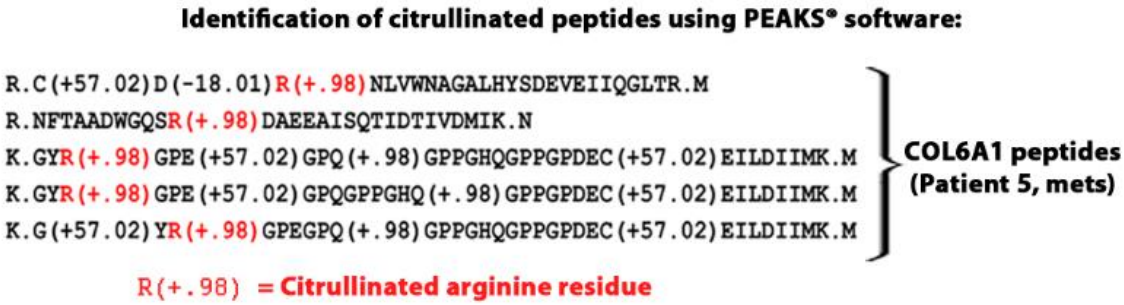


Figure 21 Identification of citrullinated peptides. PEAKS® software was utilised to quantify citrullinated peptides in the ECM of colorectal hepatic metastases. Citrullinated arginine residues are defined as R(+.98) and marked in red. For illustration, shown are the citrullinated peptides of collagen 6 in the metastasis sample of patient 5.

As seen from Figure 22, the ECM from colorectal cancer liver metastases contained an increased number of citrullinated peptides, most of which were from proteins classified as matrisomal (Figure 23). The full list of citrullinated peptides is presented in Supplementary Table 3.

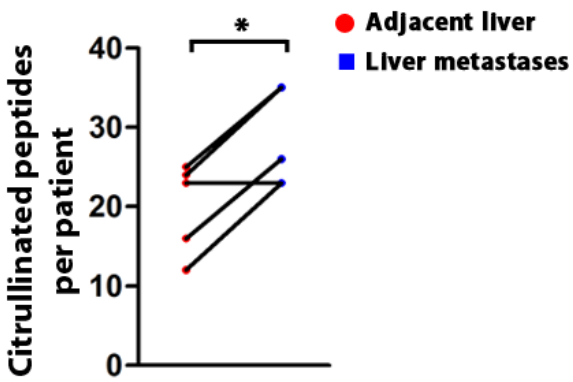


Figure 22 Comparison of citrullinated peptides between groups. Quantification of citrullinated peptides identified in the ECM from colorectal hepatic metastases and paired adjacent unaffected liver tissues (n=5 per group). Error bars indicate s.e.m., centre values indicate mean (*$P < 0.05$, two-tailed paired Student's t-test).

It is important to note that some proteins frequently citrullinated in RA, such as Vimentin, Fibrin, Tenascin, were mostly unmodified in colorectal cancer liver metastases[433]. On the other hand, proteins previously not reported as citrullinated, for

Citrullinated in adjacent liver only

Category	Protein	Uniprot	Citrullinated peptides	
			Adjacent liver	Metastasis
	HSPG2	P98160	7	0
	COL6A2	P12110	3	0
	TGM2	P21980	2	0
	ELN	P15502	1	0
	LAMB2	P55268	1	0
	COL5A1	P20908	1	0
	COL22A1	Q8NFW1	1	0
	COL6A6	A6NMZ7	1	0
	KRT6A	P02538	1	0
	MTHFD1	P11586	1	0
	ZNF469	Q96JG9	1	0
	DSEL	Q8IZU8	1	0
	TUFM	P49411	1	0
	SYNCRIP	O60506	1	0
	HNRNPA2B1	P22626	1	0
	IGHA1	P01876	1	0
	PCBP4	P57723	1	0
	FRYL	O94915	1	0
	SEC24D	O94855	1	0
	ALB	P02768	1	0

Citrullinated in both

Category	Gene name	Uniprot	Citrullinated peptides	
			Adjacent liver	Metastasis
	FBN1	P35555	8	23
	EMILIN1	Q9Y6C2	4	11
	COL1A1	P02452	6	10
	COL6A1	P12109	7	9
	FN1	P02751	4	8
	COL1A2	P08123	6	6
	COL4A2	P08572	2	5
	COL4A1	P02462	1	5
	COL2A1	P02458	6	4
	VTN	P04004	3	3
	COL5A2	P05997	2	3
	FGA	P02671	1	3
	COL6A3	P12111	4	2
	FGB	P02675	3	2
	TNC	P24821	3	2
	FUS	P35637	3	2
	COL3A1	P02461	2	2
	LAMG1	P11047	1	2
	EEF1A1	P68104	1	2
	FGG	P02679	1	1

Citrullinated in metastases only

Category	Gene name	Uniprot	Citrullinated peptides	
			Adjacent liver	Metastasis
	ACTB	P60709	0	5
	ACTA1	P68133	0	3
	PFN1	P07737	0	3
	COL8A1	P27658	0	2
	COL11A1	P12107	0	2
	REV3L	O60673	0	2
	HNRNPU	Q00839	0	2
	EFEMP1	Q12805	0	1
	EMILIN2	Q9BXX0	0	1
	FBLN5	Q9UBX5	0	1
	LTBP2	Q14767	0	1
	NID2	Q14112	0	1
	ELANE	P08246	0	1
	MUC2	Q02817	0	1
	KRT7	P08729	0	1
	PFKL	P17858	0	1
	DBNL	Q9UJU6	0	1
	HSPB1	P04792	0	1
	HSP90AA1	P07900	0	1
	OLFM4	Q6UX06	0	1

Figure 23 Citrullinated peptides identified by proteomics. Citrullinated peptides were assigned to their corresponding proteins and ranked in accordance with ECM categorization and their abundance in metastasis and adjacent liver groups. Shown are proteins identified exclusively citrullinated in metastasis, or adjacent liver, or shared by both groups.

example, Fibrillin-1 and Emilin-1, were highly citrullinated in metastatic lesions (Figure 23). This could indicate on differential roles of citrullination in RA and cancer.

To extend our results from human samples, we further examined citrullinated peptide content in the ECM from murine experimental hepatic metastases (HT29 and LoVo). There was considerable overlap with the data from human metastases, with Fibrillin-1, Emilin-1, with collagens being the most abundantly citrullinated proteins in liver metastases (Figure 24A, B). Because mass spectrometry can discriminate between human and mouse-derived proteins, we could also assess the host and tumour-specific contribution to the ECM. Consistent with results demonstrated in chapter 3, mouse-derived proteins constituted the majority of the ECM; tumour cells predominantly produced secreted factors, ECM regulators and ECM-affiliated molecules (data not shown).

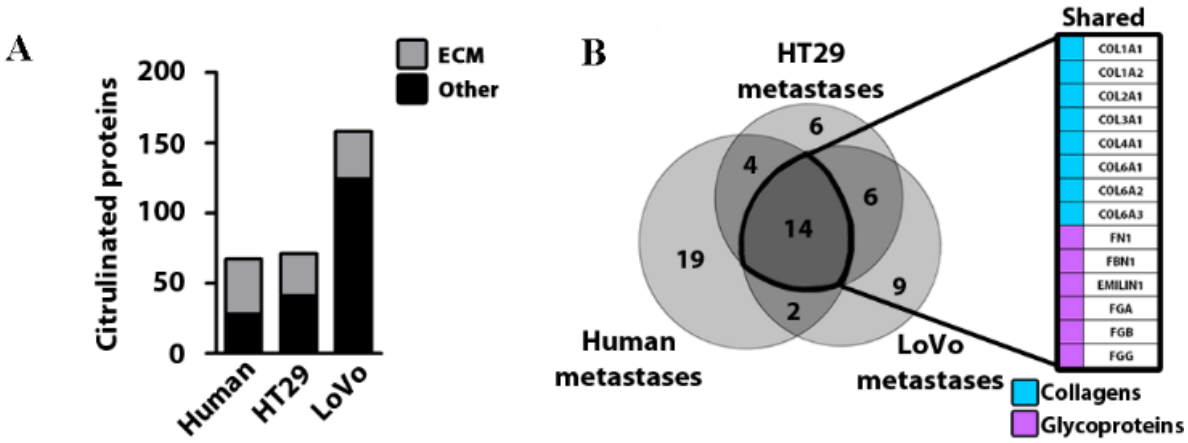


Figure 24 Analysis of ECM citrullination in mice. A. Quantification of citrullinated proteins (categorized as "ECM" or "other") identified in the ECM from human colorectal liver metastasis tissues (n=5) and ECM of experimental liver metastases generated using HT29 (n=2 biological replicates) and LoVo (n=2 biological replicates) colorectal cells. B. Venn diagram demonstrating an overlap of citrullinated ECM proteins identified in human and experimental colorectal hepatic metastases from A. Shared proteins are listed in the box to the right.

To further corroborate our findings, we then analysed the extent of overall citrullination in a different set of human hepatic metastases (*n*=12) using two different methods, an ELISA for citrulline and immunoblotting for 2, 3-butanedione monoxime modified citrulline[434]. Importantly, none of these methods detects free citrulline, which could interfere with the results. We found significantly increased citrullination in liver metastases compared to adjacent liver, primary colorectal adenocarcinomas, or normal uninvolved colon samples using both approaches (Figure 25A, B, Figure 26A, B). Taken together, these results show that citrullination of ECM proteins is a characteristic of hepatic metastases but not primary colon adenocarcinomas suggesting that citrullination could possibly be an important step in the metastatic cascade.

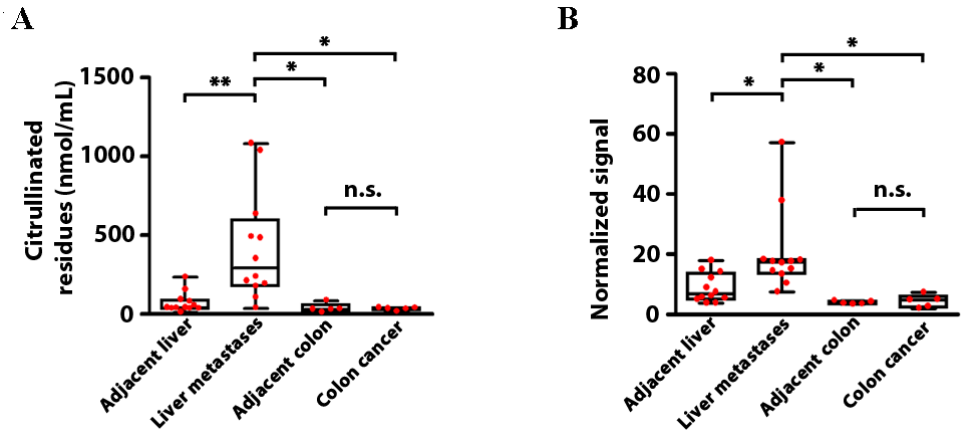


Figure 25 Analysis of citrullinated protein content in cancer and normal tissues. A. ELISA for citrullinated proteins was performed on tissue lysates from patients with colorectal hepatic metastases, paired adjacent unaffected liver tissues (*n*=12 per group), patients with primary colorectal and paired adjacent unaffected colon tissues (*n*=5 per group). Representative of two experiments. B. Densitometry analysis of immunoblotting for citrullinated proteins performed on tissue lysates extracted from patients with colorectal hepatic metastases and paired adjacent unaffected liver tissues (*n*=12 per group) and patients with primary colon adenocarcinomas and paired adjacent unaffected colon tissues (*n*=5). For the densitometry, the sum of monomer and dimer was used. Densitometry was normalised to loading control. Representative of two experiments. Error bars indicate s.e.m., centre values indicate mean (P* < 0.05, ***P* < 0.01, n.s. = non-significant, two-tailed paired Student's t-test).**

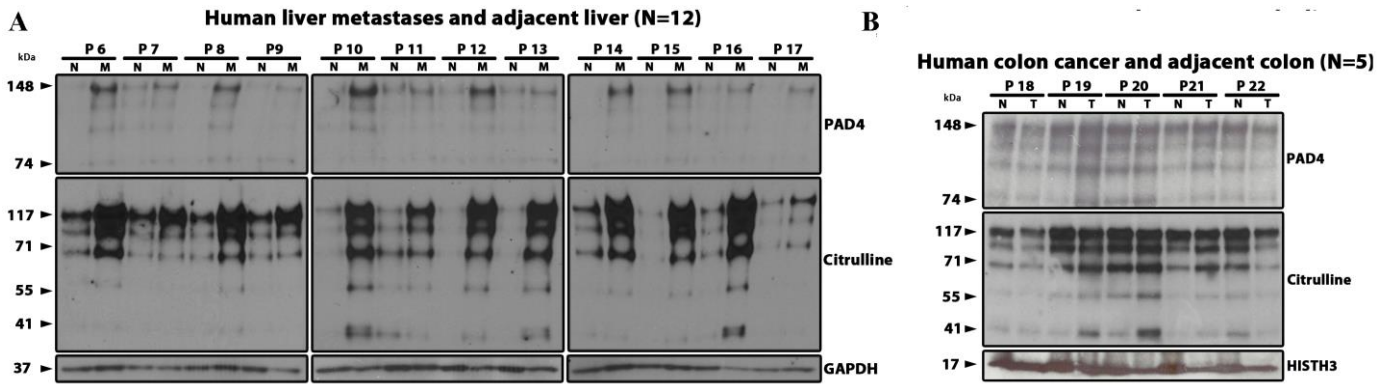


Figure 26 Immunoblotting for PAD4 and citrullinated proteins. A. Immunoblotting for PAD4 and citrullinated proteins performed on tissue lysates extracted from hepatic metastasis tissues and adjacent unaffected liver tissues (n=12 per group). Normal and metastasis tissues are labelled N and M, respectively. GAPDH was used as a loading control. B. Immunoblotting for PAD4 and citrullinated proteins performed on tissue lysates extracted from primary colon adenocarcinoma lesions and adjacent unaffected colon tissues of an independent sample of 5 patients. Normal and tumour tissues are labelled N and T, respectively. Histone H3 was used as a loading control.

4.2.1.2. *PAD4 is overexpressed in the ECM of hepatic metastases*

PAD4, a protein that catalyses citrullination, was 11 times more abundant in the metastatic ECM than the uninvolved liver, raising the possibility that this molecule was responsible for citrullination in liver metastases (Figure 10). Notably, none of the other PADs were detected in the ECM by proteomics. PAD4 levels in lysates of human liver metastases were elevated compared to the adjacent liver (Figure 26A, B, Figure 27A, B). PAD4 exists as a 74 kDa monomer, which is less enzymatically active than dimeric PAD4 [435]. Our experiments identified PAD4 in both forms whilst the dimeric form was generally more abundant (Figure 26A, B). The concentration of Ca^{+2} which regulates dimerisation is likely to vary depending on the nature of the sample[435]. We did not find increased levels of PAD4 in tissue lysates from human primary colorectal or in adjacent uninvolved colon samples, suggesting that PAD4 upregulation is specific to liver metastases (Figure 26A, B, Figure 27A, B).

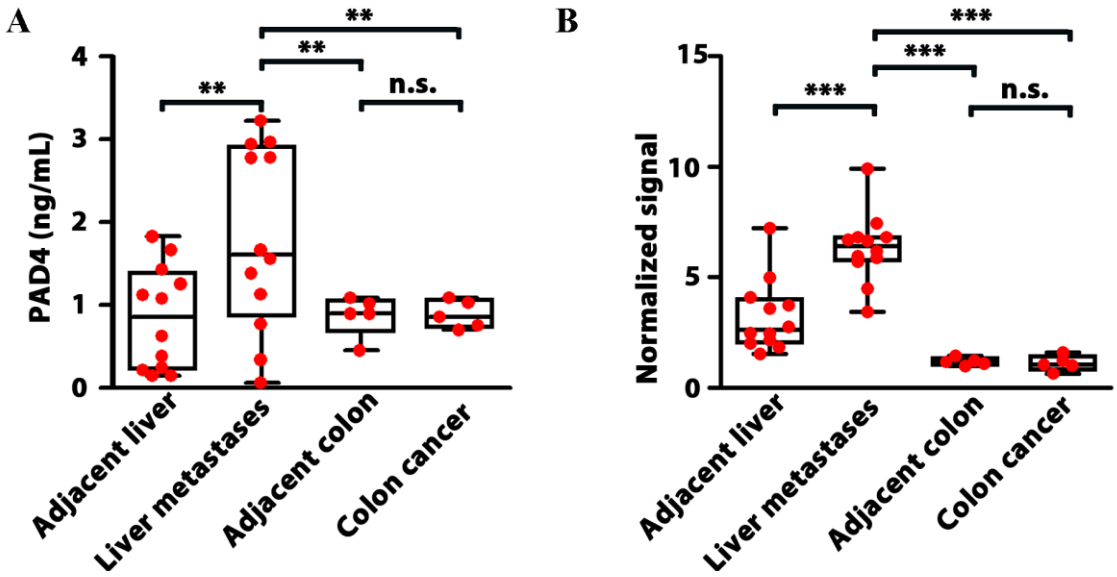


Figure 27 Analysis of PAD4 in cancer and normal tissues. A. ELISA for PAD4 was performed on tissue lysates extracted from patients with colorectal hepatic metastases, paired adjacent unaffected liver tissues (n=12 per group), patients with primary colorectal and paired adjacent unaffected colon tissues (n=5 per group). B. Densitometric analysis of immunoblotting for PAD4 performed on tissue lysates extracted from patients with colorectal hepatic metastases and paired adjacent unaffected liver tissues (n =12) and patients with primary colorectal and paired

adjacent unaffected colon tissues (n=5). For the densitometry, only the 148 kDa protein band was measured. Densitometry was normalised to loading control.

We then questioned which cell types produce PAD4 in the hepatic metastatic environment. To address this question, we sorted cells from murine experimental liver metastases generated from 3 different human colorectal cancer cell lines into 4 groups: cancer cells expressing GFP, granulocytes ($CD11b^{+}Ly6G^{hi}$), other myeloid cells ($CD11b^{+}Ly6G^{lo}$) and the remaining stromal cells (GFP^{-}, $CD11b^{-}$, $Ly6G^{-}$). Strikingly, for each tumour type, cancer cells produced considerably higher amounts of PAD4 when compared with host cells (Figure 28).

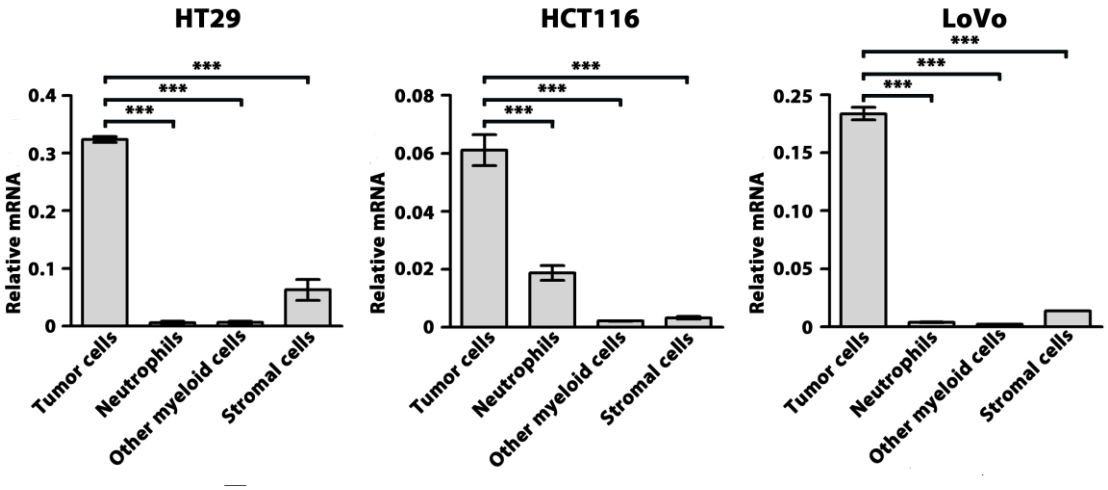


Figure 28 PAD4 profiling in FACS-sorted cells. Relative mRNA levels (normalized to HPRT) of PAD4 from cancer cells expressing GFP, granulocytes ($CD11b^{+}Ly6G^{hi}$), other myeloid cells ($CD11b^{+}Ly6G^{lo}$) and the stromal cells (GFP^{-}, $CD11b^{-}$, $Ly6G^{-}$) isolated from HT29, HCT116 and LoVo experimental hepatic metastases (n=2 biological replicates per group). Representative of two experiments.

To further confirm that neutrophils do not produce the majority of PAD4 in metastatic microenvironment, we depleted neutrophils in metastasis-harbouring mice using the systemic administration of anti-Ly6-G antibody[63] and examined if PAD4 levels

were changed (Figure 29A). Metastases depleted of neutrophils had similar levels of PAD4 and protein citrullination as control animals (Figure 29B, C).

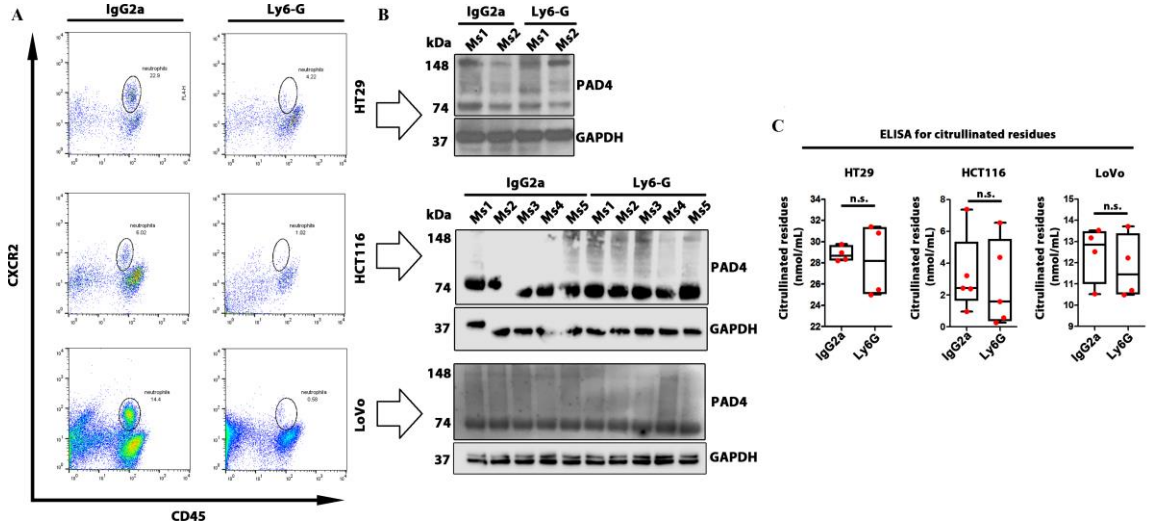


Figure 29 Neutrophil depletion does not affect citrullination in experimental liver metastases. A. Total hepatic CD45^{+}/CXCR2^{+} cell counts in liver metastasis-bearing mice treated with IgG2a or anti-Ly6G (n=5 biological replicates per group). B. Immunoblotting for PAD4 in metastasis tissue lysates (n=5 biological replicates per group). Large arrows indicate cell lines used for western blotting. GAPDH was used as a loading control. C. ELISA for citrullinated residues in metastasis tissue lysates (n=5 biological replicates per group). Error bars indicate

We found that PAD4 protein was expressed by multiple cancer cell lines, including those isolated from metastatic sources (SW62, LoVo) and it was also found in experimental mouse liver metastases (Figure 30). Collectively, the results from these experiments show that cancer cells not uncommonly express PAD4 and appear to be the predominant source of PAD4 in the metastatic microenvironment in murine models.

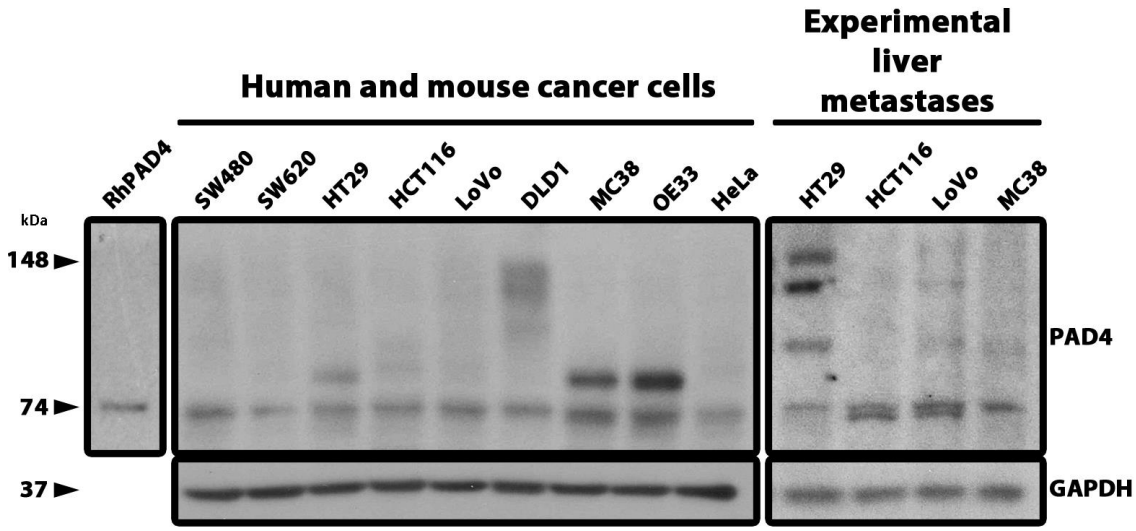


Figure 30 Immunoblotting for PAD4 performed on cell lysates of indicated human and mouse cancer cell lines and tissue lysates of experimental hepatic metastases. GAPDH was used as a loading control.

4.2.1.3. PAD4 is delivered to the ECM at least in part by exosomal transport

We then attempted to explore the mechanism through which cellular PAD4 is translocated to the ECM. PAD4 concentration increased with time in colorectal cancer cell conditioned media (Figure 31A). Notably, extracellular PAD4 was mostly present in the dimeric form, suggesting active citrullination capability outside of the cell (Figure 31B).

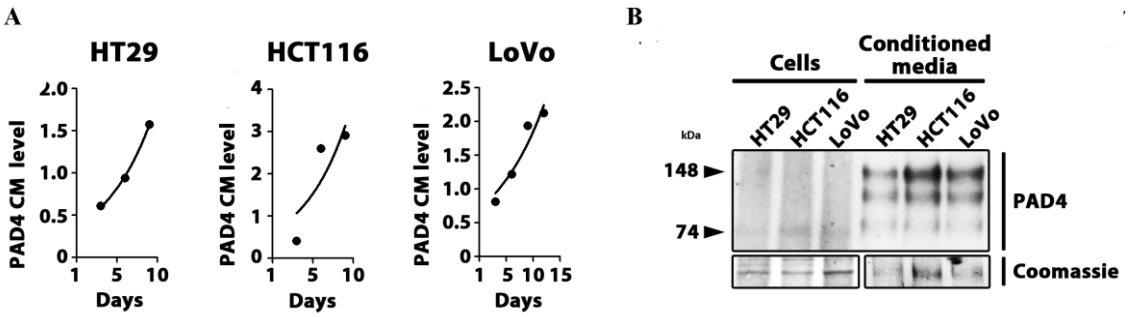


Figure 31 PAD4 accumulates in conditioned media of cancer cells. A. ELISA for PAD4 was performed on the conditioned media collected from cultured HT29, HCT116 and LoVo cells at the indicated time points. The concentration of PAD4

was normalised to the total cell number. B. Immunoblotting for PAD4 was performed on HT29, HCT116 and LoVo cells and their corresponding conditioned media after 6 days of culture. Coomassie stain was used as a loading control.

Although this data demonstrates secretion of PAD4 by colorectal cells, PAD4 lacks a defined secretory peptide raising the possibility that it is secreted in cellular vesicles such as exosomes. This mechanism would be consistent with the extensive deposition of proteins associated with exosomes in the ECM (Figure 8). In keeping with deposition of PAD4 through exosomal release, exosomes isolated from cancer cell conditioned media (Figure 32A) contained PAD4 as well as the characteristic exosomal markers caveolin-1 and β-catenin (Figure 32B).

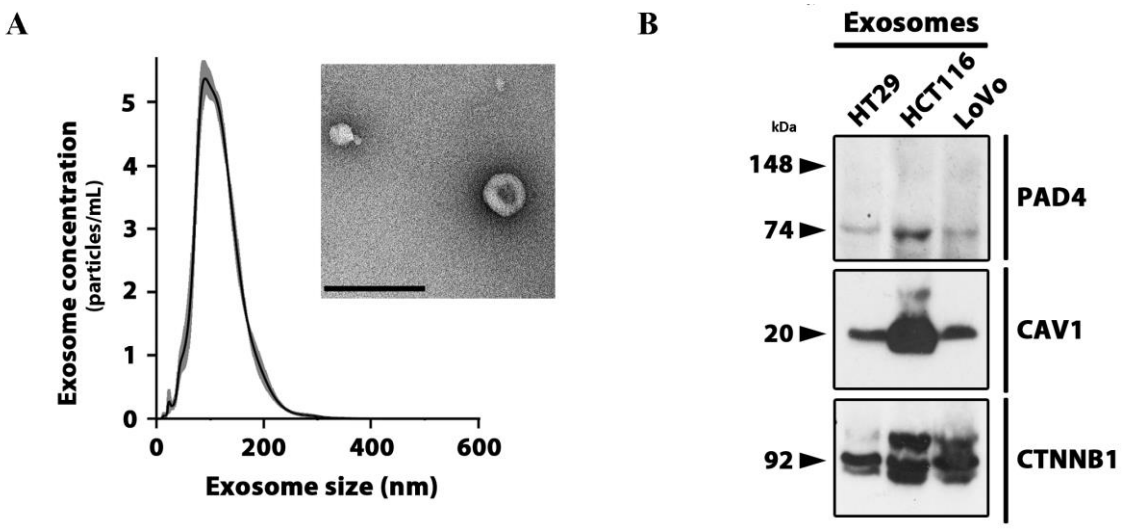


Figure 32 Exosomes containing PAD4. A. Dynamic light scattering analysis of particle size distribution (graph) and transmission electron microscopy (image) of exosomes isolated from cultured HT29 cells. Scale bar = 250 nm. B. Immunoblotting for PAD4, CAV1 and CTNNB1 in exosomes isolated from cultured HT29, HCT116 and LoVo cells. Representative of two experiments.

Exposure of colorectal cancer cells to the sphingomyelinase inhibitor GW4869, a potent inhibitor of exosome release[436], led to more than a two-fold reduction in exosome number (Figure 33A), and a corresponding decrease in PAD4 levels in the conditioned

media of HT29 colorectal cancer cells, without altering intracellular PAD4 levels (Figure 33B). GW4869 did not affect the proliferation of HT29 cells; however, we did not test GW4869 on HCT116 and LoVo cells due to toxicity (data not shown).

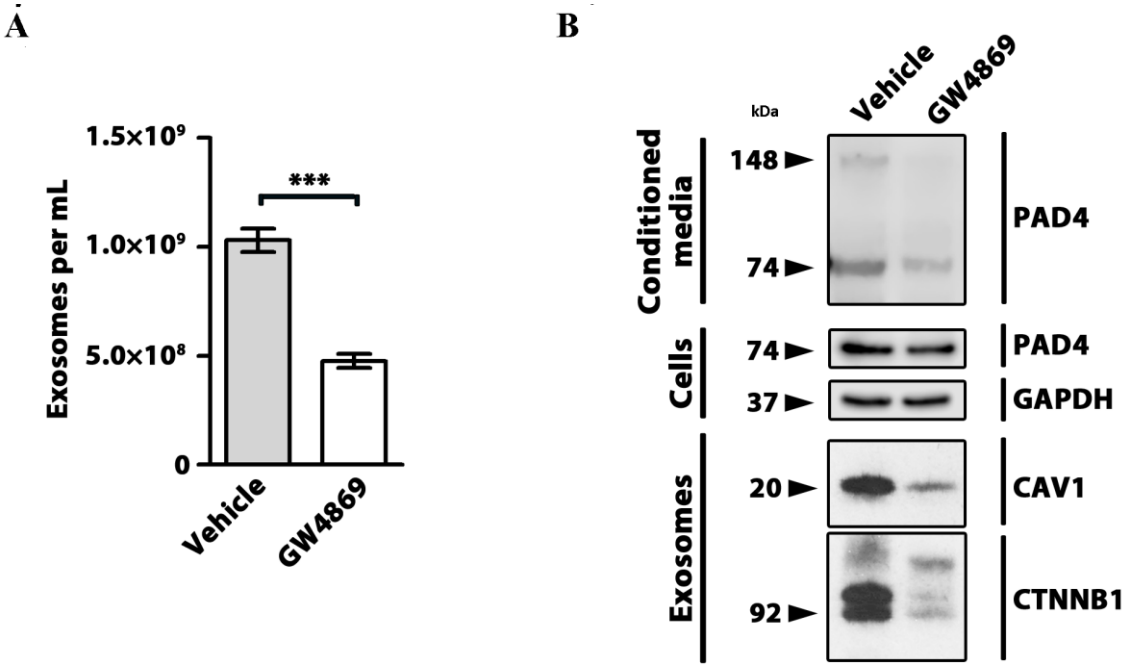


Figure 33 Exosome inhibition using GW4869. A. Dynamic light scattering analysis of particle concentration of exosomes collected from HT29 cells treated with vehicle or GW4869 for 3 days. Representative of two experiments. B. Immunoblotting for PAD4, CAV1 and CTNNB1 in cells, conditioned media and exosomes from cultured HT29 cells treated with vehicle or GW4869 for 3 days. GAPDH was used as a loading control. Representative of two experiments.

Finally, immunohistochemistry of human colorectal cancer hepatic metastases revealed both intracellular and extracellular PAD4 staining. The extracellular PAD4 colocalised to collagen type I providing additional evidence for the presence of PAD4 in the tumour ECM (Figure 34). Of note, secretion to the extracellular space of proteins without secretory peptides such as IL-1b has previously been documented[437]. Whilst this data does not exclude the possibility that PAD4 might be released through cancer cell

death, it suggests that PAD4 is derived from cancer cells and can be delivered to the ECM via exosomes.

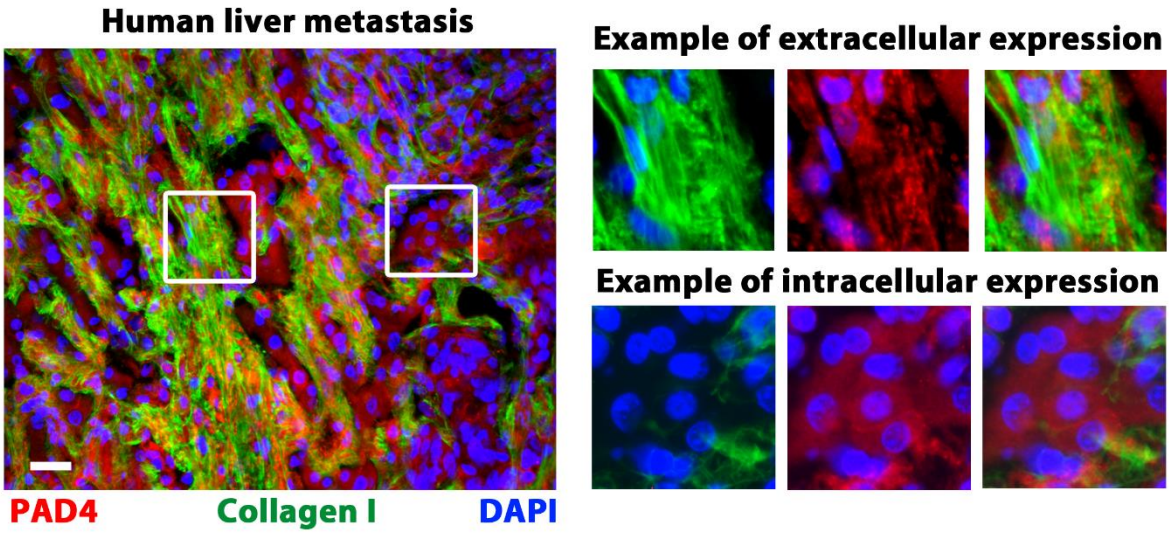


Figure 34 Representative image of co-immunostaining for PAD4 and collagen I in human colorectal hepatic metastases tissue. Scale bar = 100 µm.

4.2.2. The function of PAD4/Citrullination in hepatic metastases

4.2.2.1. Plating cancer cells on citrullinated collagen type I enhances cancer cell adhesion, reduces motility and alters epithelial-mesenchymal plasticity

In order to examine the impact of ECM citrullination on phenotype of tumour cells, we studied colorectal cancer cells plated on citrullinated and non-citrullinated collagen type I. We chose specifically collagen type I for these studies because it was one of the most highly expressed component of the liver metastatic ECM and because it was extensively citrullinated in human liver metastases (Figure 23). To generate citrullinated collagen, we incubated recombinant collagen type I with recombinant PAD4, as assessed by mass spectrometry analysis (Figure 35). As a control we added the PAD inhibitor BB-Cl-amidine[205] to inhibit citrullination (Figure 35).

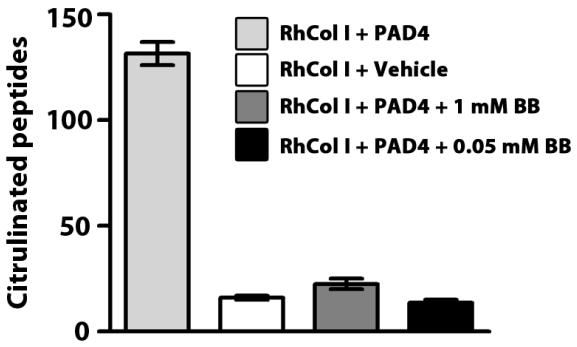


Figure 35 Recombinant collagen type I alone, collagen type I pre-treated with recombinant PAD4, or collagen I pre-treated with recombinant PAD4 and BB-Cl-amidine, were subjected to LC-MS/MS analysis as described in methods. Briefly, recombinant collagen I (ab6308; Abcam) was incubated at a concentration of 40 µg/mL with 200 ng active human PAD4 (ab196393; Abcam) in the presence or absence of 1 mM or 0.05 mM BB-Cl-amidine for 18 h at 37ºC in a buffer of 20 mM HEPES (pH 8.8), 10 mM CaCl2, 1 mM DTT, 1 mM EDTA, and 0.3 M NaCl. The concentrations of PAD4 and collagen I were used to obtain the enzyme-to-substrate ratio of 1:10, as previously described[255]. The reaction was stopped by the addition of EDTA to a final concentration of 10 mM. Samples were analysed by LC-MS/MS for the presence of citrulline residues. Shown is the quantification of citrullinated peptides. Data is from two independent experiments.

Adhesion of colorectal cancer cells was significantly greater to citrullinated collagen type I than to the non-citrullinated controls (Figure 36A, B). Cell motility, measured by median velocity and track length was also significantly decreased in colon cancer cells plated on citrullinated collagen type I compared to controls (Figure 37A, B). α1β1 and α2β1 are the most prevalent collagen binding integrins; blocking antibodies to these integrins abrogated cancer cell binding to both citrullinated and non-citrullinated collagen type I, indicating that observed changes in adhesion are integrin dependent (Figure 36B). These data implicate citrullination of the ECM in the promotion of cancer cell adhesion and inhibition of their motility, characteristic features of an epithelial cell phenotype.

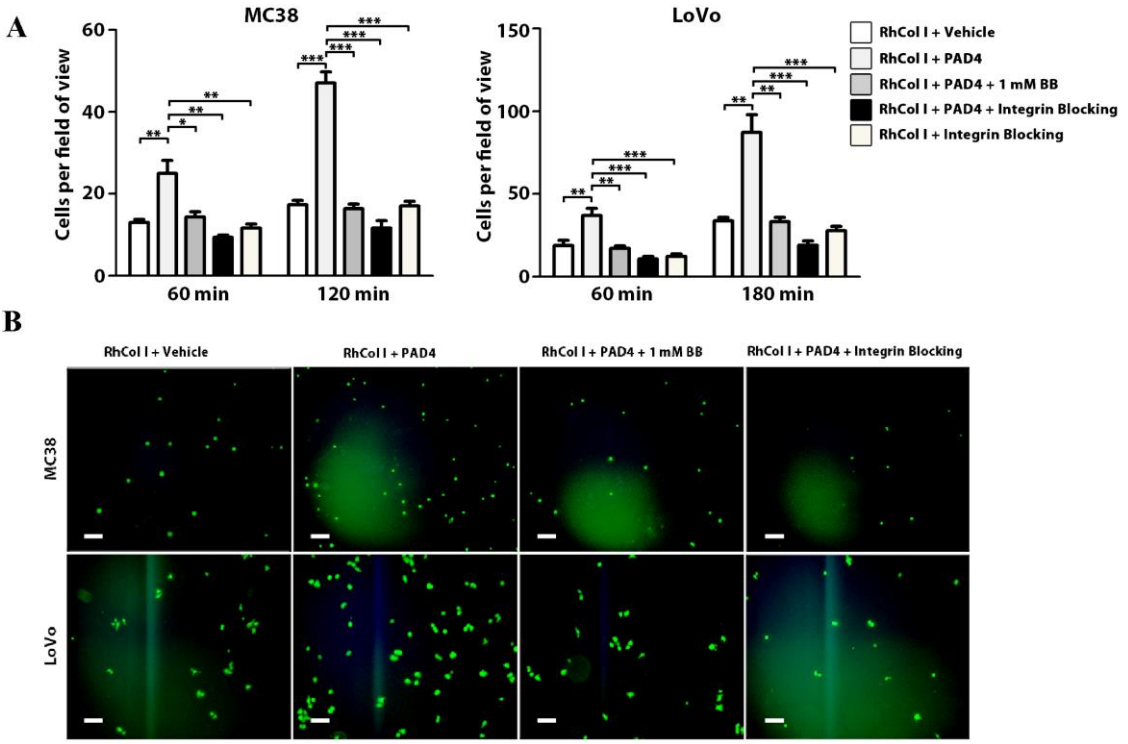


Figure 36 Analysis of cancer cell adhesion. A. GFP$^+$ cancer cells, MC38 or LoVo as indicated, were plated in wells pre-coated with either collagen type I alone, collagen type I pre-treated with recombinant PAD4, collagen type I pre-treated with PAD4 and BB-Cl-amidine, collagen I pre-treated with PAD4 and integrin blocking antibodies, or collagen I and integrin blocking antibodies alone. Cells were imaged at the indicated time points using an epifluorescence microscope. At least 5 fields of view per condition were taken using 20x objective. Shown are the cell numbers attached to the plate for each condition. Data is from two experiments. B. Representative images from A. Scale bar = 100 µm. For A, error bars indicate s.e.m. centre values indicate mean (*P < 0.05, **P < 0.01, ***P < 0.001, two-tailed unpaired Student's t-test).

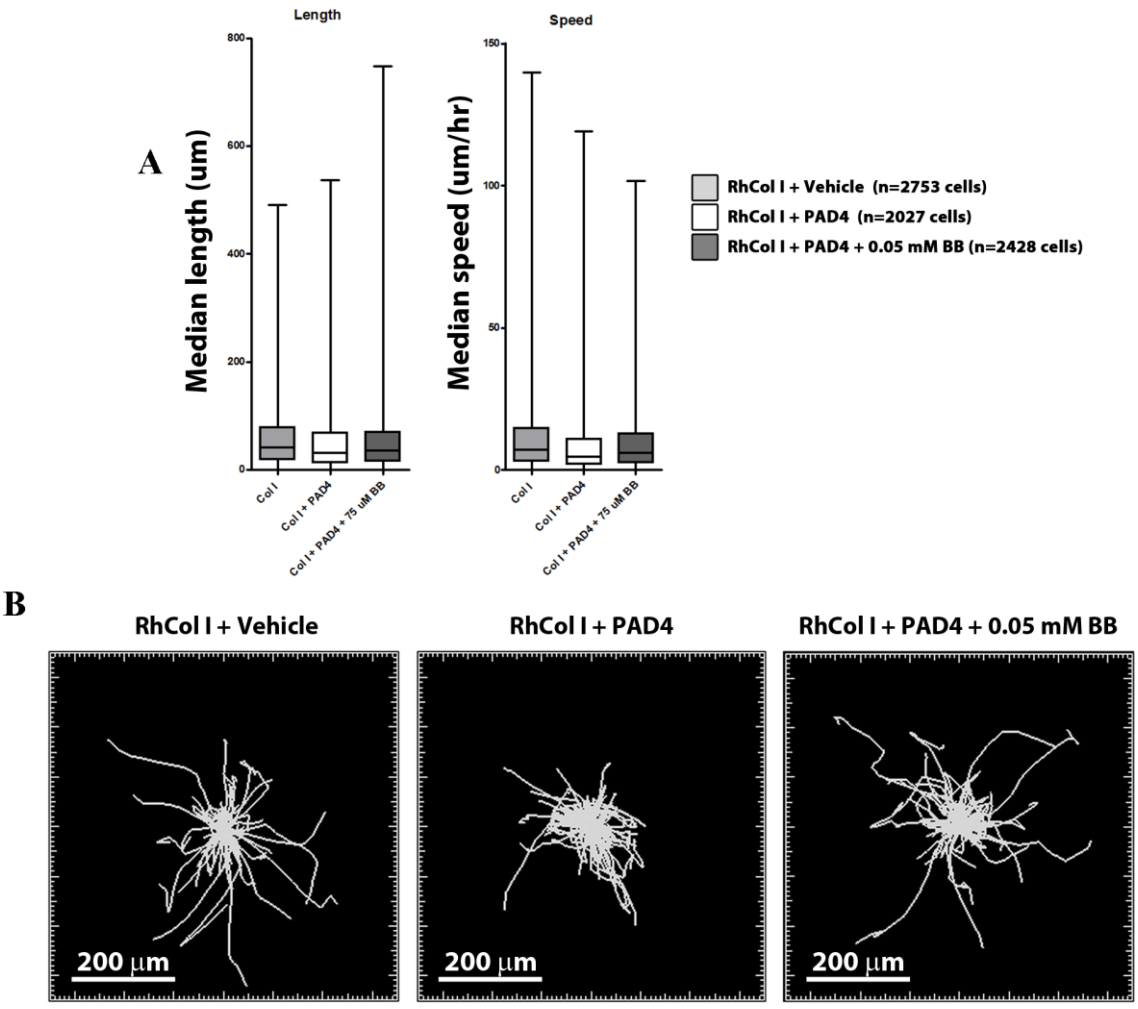


Figure 37 Analysis of cancer cell motility. A. GFP^{+} MC38 cells were plated in wells pre-coated with either recombinant collagen type I alone, recombinant collagen type I pre-treated with recombinant PAD4, or collagen type I pre-treated with addition of PAD4 and BB-Cl-amidine. Individual cells were tracked using time-lapse microscopy. Shown is the quantification of median length (left) and speed (right) of individually-tracked cells. Representation of two experiments. B. Representative images of cell tracks from the experiment in A. For A, error bars indicate interquartile range, centre values indicate median (*P < 0.05, ***P < 0.001, Kruskal-Wallis test with Dunn`s multiple comparison post-test).

We then examined the effect of adhesion to citrullinated collagen type I on downstream signaling mediated by integrin binding to collagen type I. Focal adhesion kinase (FAK) is an integrin-dependent regulator of cell motility linked to EMT[438]. Phosphorylation of FAK and of its downstream targets, extracellular signal-regulated kinase (ERK) and JNK, was decreased in colorectal cancer cells seeded onto citrullinated

collagen type I in comparison to control, consistent with the decreased motility observed (Figure 38A). Further, when plated on citrullinated collagen type I, the highly invasive MC38 colorectal cancer cells exhibited epithelial markers perhaps indicative of MET phenotype characteristic of the reverse of EMT. In particular, mRNA expression of mesenchymal markers *N-cadherin* (*CDH2*) and *Snail1* and *Snail3* (*SNAI1* and *SNAI3*) were reduced, whereas the expression levels of *E-cadherin* (*CDH1*) and other epithelial markers were increased (Figure 38B). Reciprocal changes in *N-cadherin* and *E-cadherin* were further validated at the protein level by immunoblotting (Figure 38C). Taken together, these data show that adhesion to citrullinated collagen type I alters the phenotype of parameters associated with EMT in colorectal cell lines.

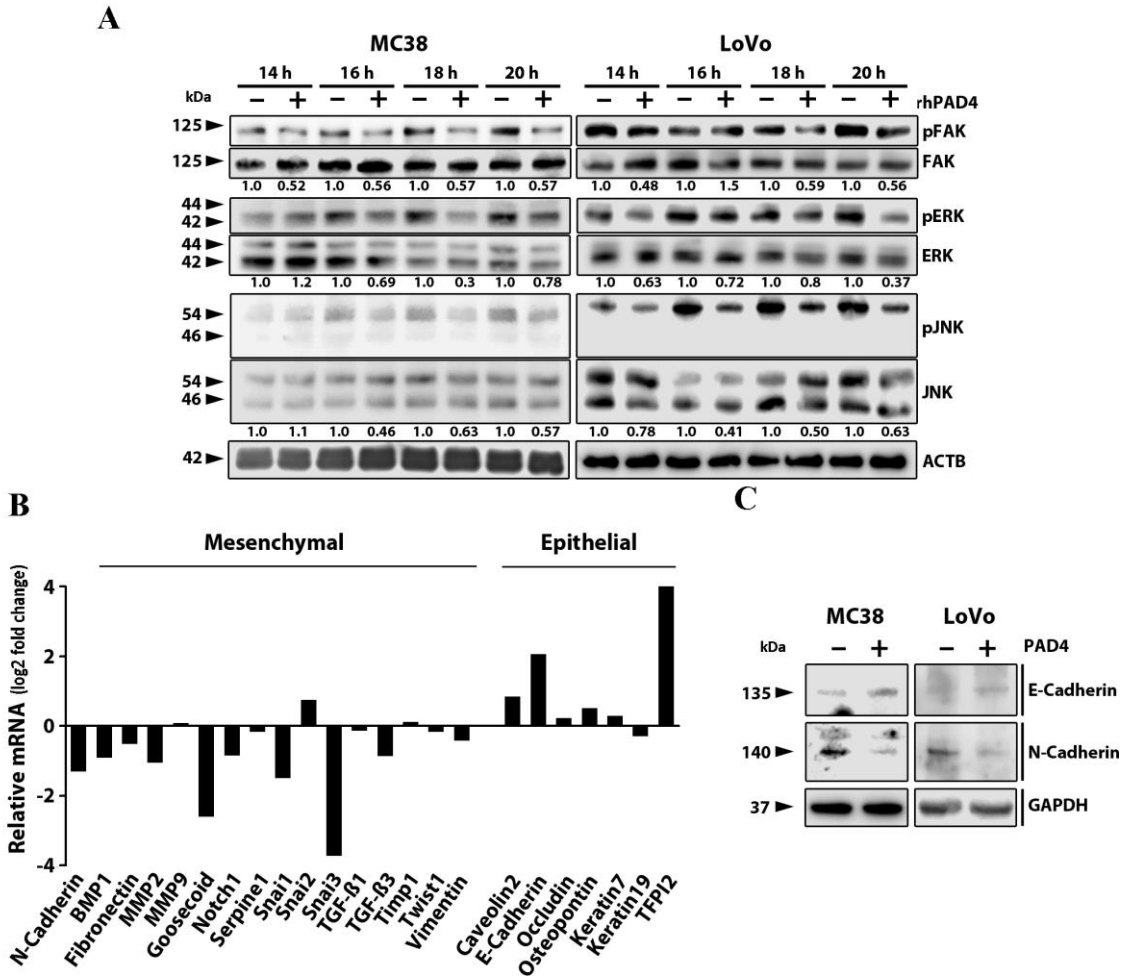


Figure 38 Analysis of FAK pathway and EMT factors. A. Cancer cells were plated in wells pre-coated with either recombinant collagen type I alone or collagen type I pre-treated with recombinant PAD4 (+rhPAD4). At indicated times cells were

collected and probed for the indicated proteins. Beta-actin was used as a loading control. Numbers indicate relative expression of a phosphorylated protein based on densitometry analysis of protein bands normalised to the pixel density of non-phosphorylated protein and loading control. B. Shown is relative mRNA expression of EMT genes (normalised to HPRT) in MC38 cells seeded on collagen type I pre-treated with PAD4 as compared to those on control collagen. C. Immunoblotting for the indicated proteins performed on cell lysates collected from the experiment in B. GAPDH was used as a loading control. Experiments were performed twice.

4.2.2.2. Inhibition of PAD reduced liver metastatic growth and reversed MET in vivo

To define whether analogous events occur *in vivo*, we treated mice in an experimental liver metastasis experiment with the PAD inhibitor BB-Cl-amidine. This resulted in a consistent decrease in the amount of citrullinated proteins in the liver colonies of the human colorectal cells, although there was substantial variation between individual mice (Figure 39A, B, E). Because the commercial antibody to 2, 3-butanedione monoxime-modified citrulline is human-specific, we used immunohistochemistry for citrullinated histone H3 to assess the extent of citrullination in the MC38 murine experimental metastasis model (Figure 39C-E).

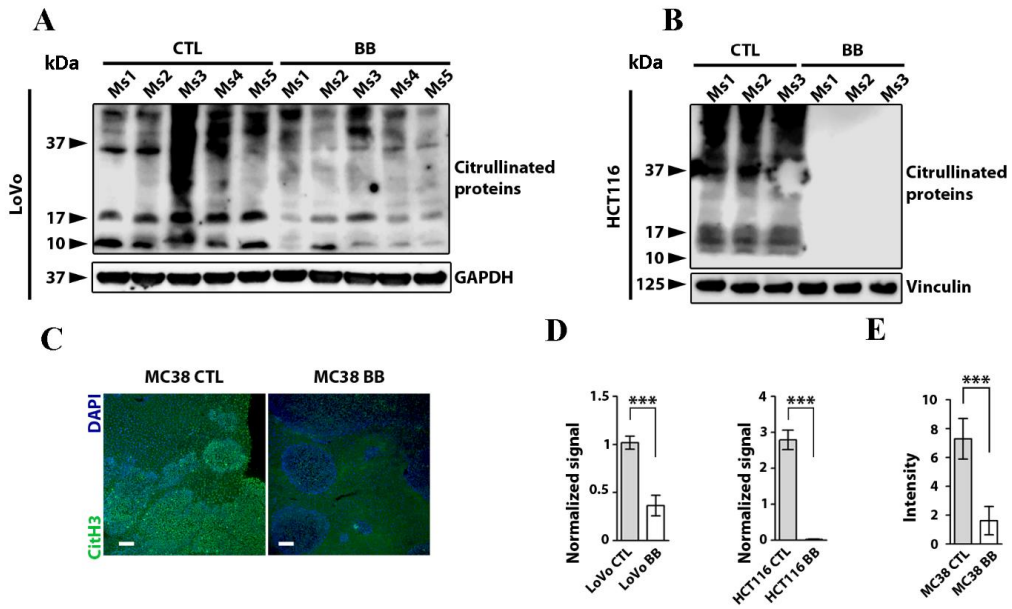


Figure 39 BB-Cl-amidine reduces citrullination in murine liver metastases. A, B. Immunoblotting for citrullinated proteins performed on tissue lysates extracted from experimental LoVo and HCT116 hepatic metastases from mice treated with vehicle or BB-Cl-amidine. Each lane indicates a biological replicate. Vinculin was used as a loading control. C. Representative images of immunostaining for citrullinated histone H3 in experimental MC38 hepatic metastases from mice treated with vehicle (n=5 biological replicates), or BB-Cl-amidine (n=6 biological replicates). n=3 sections per group, at least 4 images were analysed per section. Scale bar = 100 µm. D. Densitometric analysis for A and B normalised to loading control. E. Staining intensity quantification for C. Throughout, error bars indicate s.e.m. centre values indicate mean (***P*** **< 0.01, ******P*** **< 0.001, two-tailed unpaired Student's t-test). Experiments were performed once.**

The decrease in citrullinated proteins in the liver colonies correlated with a 3 to 4-fold decrease in metastatic burden (Figure 40A, B). Similar to our cell culture results, tumours from untreated mice with more extensive citrullination exhibited strong expression of the epithelial markers E-cadherin, Cytokeratin 7 and Tight junction protein-1 and decreased expression of the mesenchymal markers ZEB1 and N-Cadherin (Figure 40C, D). In contrast, BB-Cl-amidine-treated tumours exhibited the reverse of this, with decreased expression of epithelial markers and increased expression of mesenchymal

markers. Thus, reduction of PAD activity and decreased citrullination reverses MET in experimental liver metastases, concurrently impairing metastatic growth.

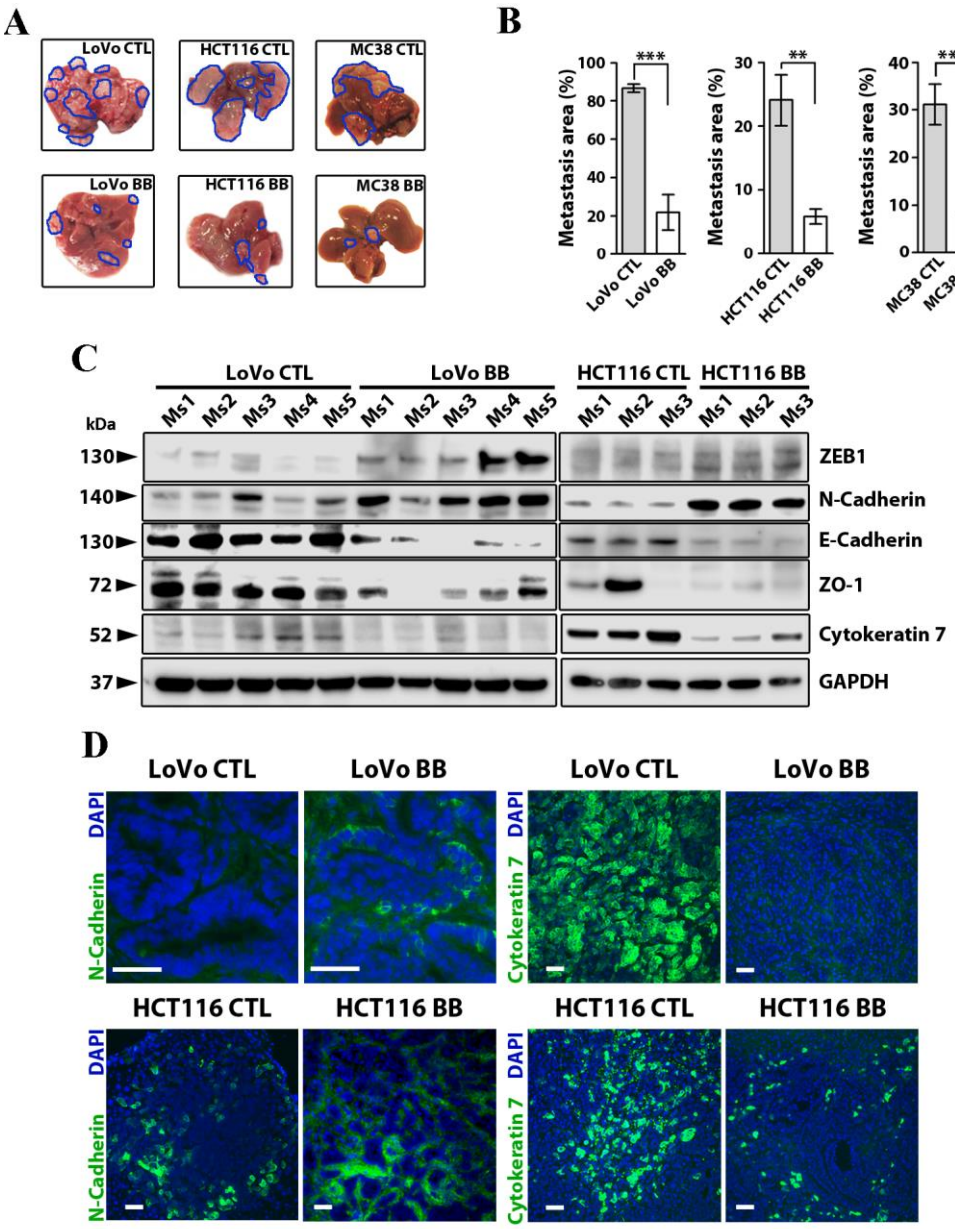


Figure 40 BB-Cl-amidine alters EMT markers in murine liver metastases. A. Representative images of livers with experimental LoVo, HCT116 and MC38 hepatic metastases from mice treated with vehicle (n=7, 6, and 5 biological replicates, respectively) or BB-Cl-amidine (n=8, 6, and 6 biological replicates, respectively). Metastatic nodules are outlined in blue. Representative of two experiments. B. Quantification of metastatic area percentage based on the assessment of scanned H&E-stained sections of livers from A. At least 3 different liver scans per mouse were analysed. C. Immunoblotting for the indicated proteins performed on tissue lysates from the experiment in A. GAPDH was used as a loading control. D. Representative images of staining for indicated proteins in experimental hepatic metastases from A. At least 5 biological replicates were

stained per group. Scale bar = 100 µm. Throughout, error bars indicate s.e.m. centre values indicate mean (***P*** **< 0.01, ******P*** **< 0.001, two-tailed unpaired Student's t-test). Experiments were performed once.**

We then assessed the effect of BB-Cl-amidine on subcutaneous tumour growth, observing approximately a 2-3-fold decrease in LoVo and MC38 xenograft volume, but no decreased growth of HCT116 xenografts (Figure 41A). Interestingly, the epithelial-mesenchymal plasticity of the subcutaneous tumours was not altered upon BB-Cl-amidine treatment, suggesting that PAD/citrullination-mediated EMT changes mainly impact upon growth after metastatic dissemination of the cancer cells (Figure 41B). Possibly, in these subcutaneous tumours the cells may already be tilted closer to the mesenchymal phenotype. Another explanation could be that we observed little citrullination/PAD4 levels in primary colon adenocarcinomas, suggesting irrelevance of this axis in primary tumours.

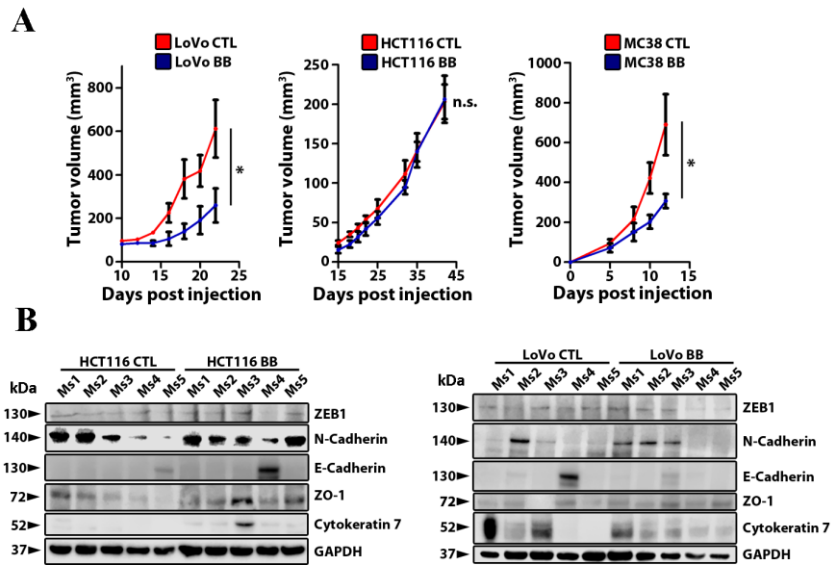


Figure 41 Effect of BB-Cl-amidine on subcutaneous xenografts. A. Tumour growth curves of mice subcutaneously injected with indicated cells and treated with vehicle (n=5, 5, and 6 biological replicates for LoVo, HCT116 and MC38 cancer cell lines, respectively) or BB-Cl-amidine (n=5, 5 and 5 biological replicates for LoVo, HCT116 and MC38 cancer cell lines, respectively). B. Immunoblotting for the indicated proteins performed on tissue lysates from A. For A, error bars indicate s.e.m. (*$P < 0.05$, ***$P < 0.01$, n.s. = non-significant, two-way ANOVA where factors were treatment group and time, the test for interaction was performed ($P < 0.05$ for LoVo and MC38, n.s. for HCT116). Experiments were performed 2 times.

We then asked whether downregulation of PAD4 in colorectal cancer cells had analogous effects to pharmacological inhibition. PAD4 levels were reduced by introduction of the shRNA in HT29 and HCT116 cells and resulted in decreased intracellular and secreted protein citrullination (Figure 42A, B). Expression of the EMT markers N- and E-cadherin, however, was unchanged suggesting that intracellular PAD4 and intracellular citrullination have minimal effects on epithelial-mesenchymal plasticity in these cells (Figure 42C). This is in striking contrast to the shift to MET observed after plating these cells on citrullinated collagen (Figure 42C). Thus, intracellular PAD4/citrullination do not account for the effects on epithelial-mesenchymal plasticity.

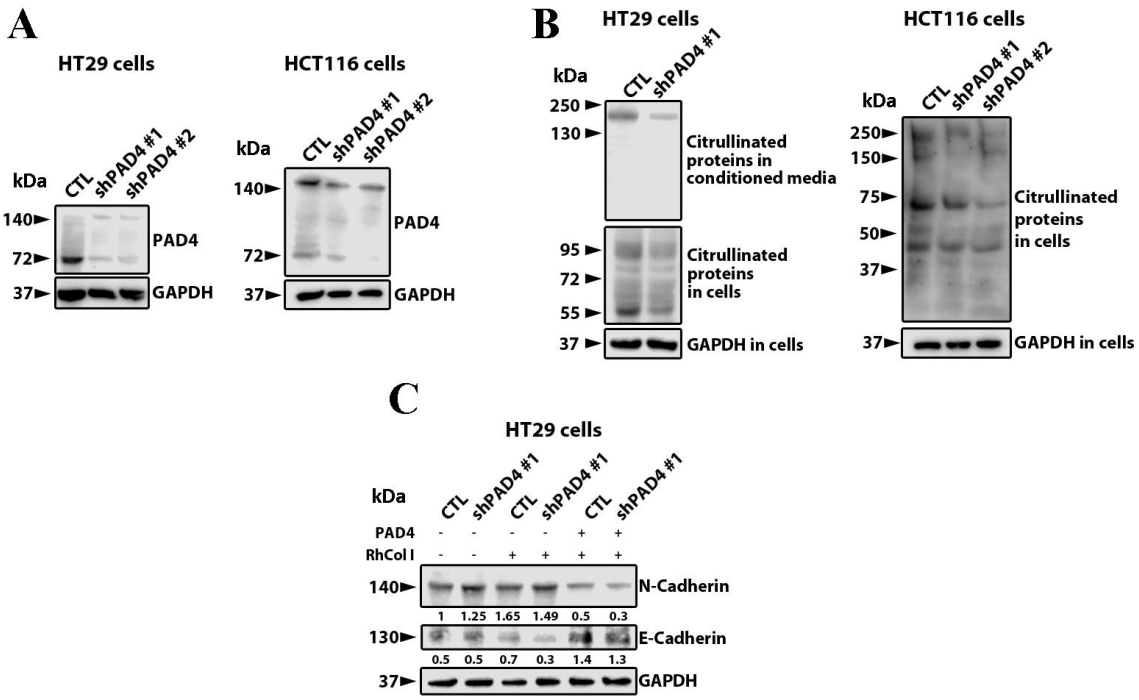


Figure 42 Immunoblotting analysis of PAD4-deficient colorectal cancer cells. A. Immunoblotting for PAD4 in HT29 and HCT116 cells transfected with empty vector (CTL) or shPAD4 lentivirus particles (shPAD4 #1 and #2). GAPDH was used as a loading control. B. Immunoblotting for citrullinated proteins in the conditioned media and cells from control and PAD4-deficient HT29 and HCT115 cells. C. Control and PAD4-deficient HT29 cells were plated on recombinant collagen type I that had been exposed to PAD4 as indicated. Cells were collected and probed for the indicated proteins. GAPDH was used as a loading control. Numbers indicate relative expression based on densitometry analysis. Experiments were performed 2 times.

The growth of PAD4-depleted HT29 and HCT116 cells was reduced in culture and in subcutaneous xenografts (Figure 43A, B). Cell lysates as well as decellularised and enriched matrices from subcutaneous xenografts demonstrated a decrease in both PAD4 and citrullinated protein levels (Figure 43C). We then injected wild-type and PAD4 knockdown cells intrasplenically to assess their growth as hepatic metastases. Strikingly, PAD4-deficient cells failed to form experimental liver metastases, suggesting the importance of this protein in the formation of liver colonies *in vivo* (Figure 43D, E).

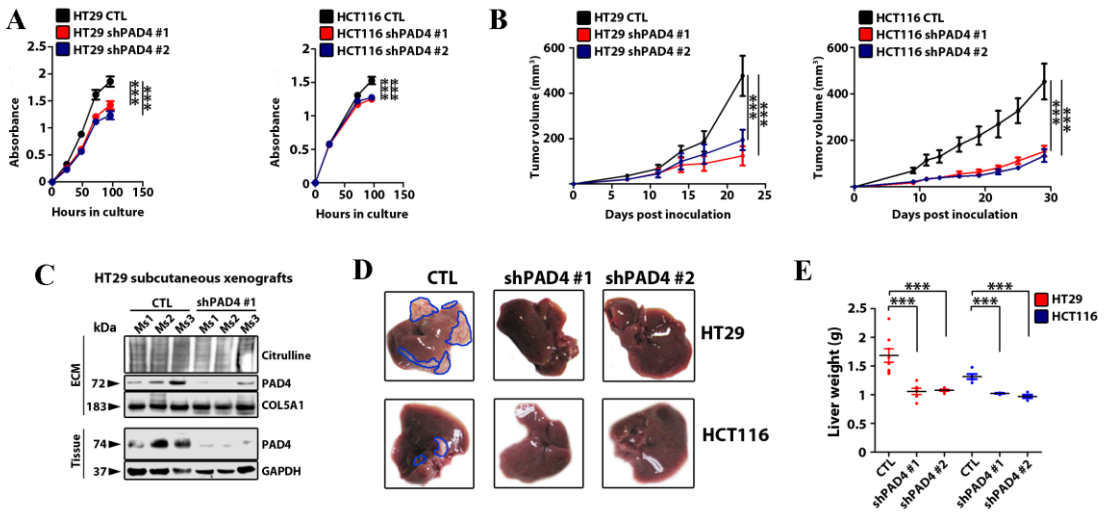


Figure 43 Effect of PAD4 knockdown on cancer cell growth *in vitro* and *in vivo*. A. WST-1 proliferation assay performed in cultured wild-type and PAD4-deficient HT29 and HCT116 cells (*n*=5 technical replicates per group). B. Tumour growth curves of mice injected subcutaneously with control and PAD4-knockdown HT29 and HCT116 cells (*n*=7, 4, and 5 biological replicates for HT29 and 9, 5, and 5, for HCT116 respectively). C. Immunoblotting analysis for PAD4 and citrulline in tissue lysates and isolated ECM from control (CTL) and PAD4 knockdown (shPAD#1) subcutaneous xenografts. GAPDH and COL5A1 were used as loading controls for tissue lysates and isolated ECM, respectively. D. Representative images of livers from mice injected intrasplenically with control (*n*=8 biological replicates) or PAD4-deficient HT29 and HCT116 cells (*n*=6 and 6 biological replicates for shPAD4 #1 and #2) 35 days post-injection. Metastatic nodules are outlined in blue. E. Liver weights from the experiment in D. Throughout, error bars indicate s.e.m. centre values indicate mean (***P*** < 0.01, ******P*** < 0.001, two-tailed unpaired Student's t-test). For A and B, error bars indicate s.e.m. centre values indicate mean (******P*** < 0.001, two-way ANOVA). Experiments were performed 2 times.**

To reinforce our findings that PAD4-deficient cells are unable to form metastases, we next performed a spontaneous metastasis assay using orthotopic injection of HCT116 cancer cells into the caecal wall of mice (Figure 44A). The resulting bowel tumours had similar weights in both wild-type and PAD4 knockdown groups (Figure 44B); however, the number of visible liver macrometastases (1-3 mm) was considerably decreased in the PAD4 knockdown group (Figure 44C). To evaluate the number of micrometastases, we sectioned the livers and counted the number of lesions as assessed by GFP positivity or histological evaluation (Figure 44D). Both methods confirmed the marked reduction of

micrometastases in the PAD4 knockdown group as compared to controls (Figure 44E, F). Unexpectedly, the expression of PAD4 was higher in metastases from mice injected with PAD4-deficient cells (Figure 44G), even though PAD4 levels were lower in the primary lesions of the PAD4 knockdown group (Figure 44H). These findings imply that the few established metastases that developed were from cells which escaped PAD4 knockdown. This result is consistent with a metastasis-specific requirement for PAD4. Taken together, these data suggest that *in vivo* activity of PAD4 in liver metastatic colorectal cancer cells drives MET and reduction of that activity diminishes growth of liver colonies.

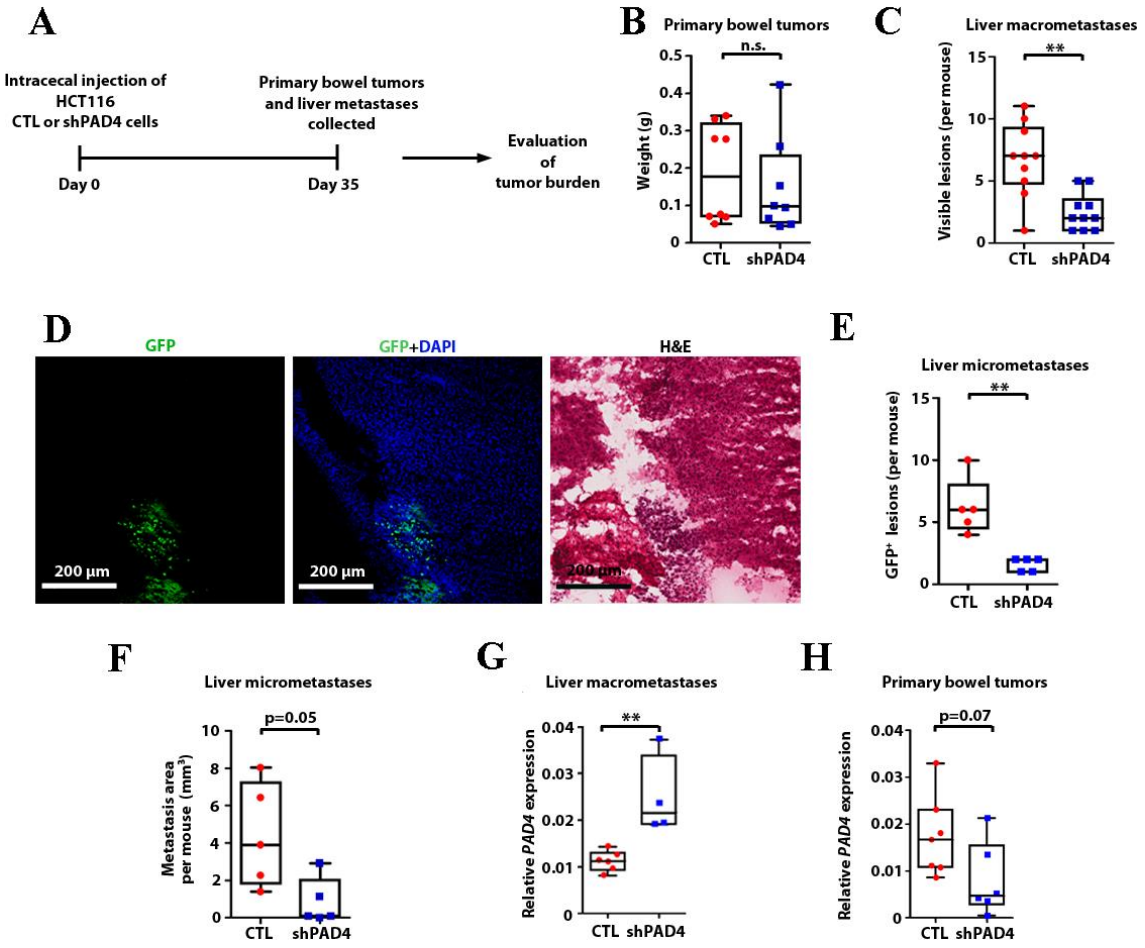


Figure 44 Effect of PAD4 knockdown on development of spontaneous metastases. A. Experimental pipeline. B. Weight of excised primary caecal tumours (n=8 biological replicates per group). C. Number of visible macrometastases (n=10 biological replicates per group). D. Representative image of a GFP$^+$ metastatic lesion counterstained with DAPI and the corresponding H&E image. E. Numbers of GFP$^+$ micrometastatic lesions (n=5 biological replicates per group). F. Area of micrometastases evaluated by H&E (n=5 biological replicates per group). G. Relative mRNA levels (normalised to HPRT) of *PAD4* in visible macrometastases (n=6 and 4 for CTL and shPAD4 groups, respectively) H. Relative mRNA levels of *PAD4* (normalised to HPRT) in primary caecal lesions (n=7 and 6 for CTL and shPAD4 groups, respectively). Throughout, whiskers indicate range, centre values indicate median (***P*** **< 0.01, two-tailed unpaired Student's t-test, n.s. = non-significant). Experiments were performed once.**

4.2.2.3. Effect of BB-Cl-amidine on tumour cell survival

We then questioned whether BB-Cl-amidine affects survival of cultured cancer cells. We have used a well-recognised method of colony growth assay with an ascending concentration of the drug. As seen from Figure 45, 1 µm of BB-Cl-amidine was enough to eliminate 50% of colonies of HT29, HCT116, LoVo and MC38 colorectal cancer cell lines. However, the mouse lymphatic endothelium cell line 2H11 did not respond to BB-Cl-amidine treatment even at 5 µm, suggesting that this compound does not affect normal cells, specifically inhibiting the survival of colorectal cancer cells.

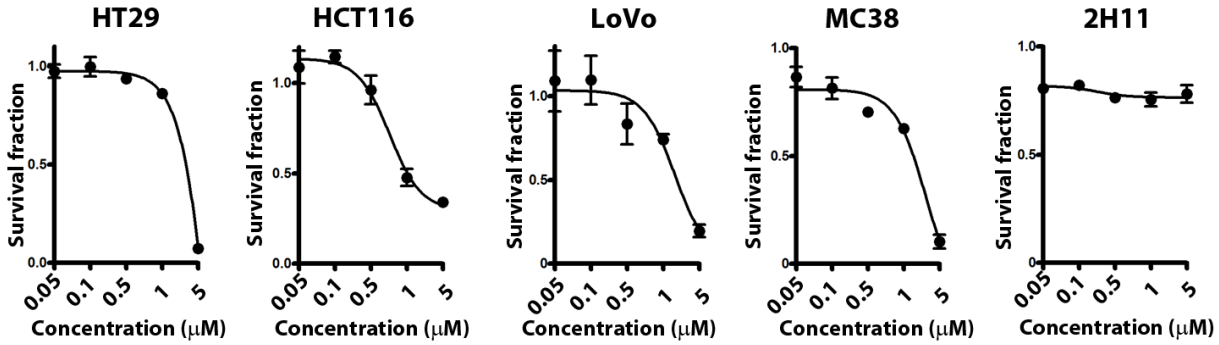


Figure 45 Colony growth assay of the indicated cell lines in response to BB-Cl-amidine. Activity of BB-Cl-amidine limits the growth of colon cancer cells yet does not affect normal lymphatic endothelium cells. Error bars represent technical replicates (n=6 per condition). Experiments were performed 2 times.

4.3. Discussion

Modifications of the ECM orchestrate many facets of cancer biology including tumourigenesis, invasion, tumour cell dissemination and metastasis[161,439,440]. In this chapter we identify citrullination generated by tumour cell-derived PAD4 as an important contributor of human colorectal metastases. The liver is the most common site for distant metastasis in colorectal cancer and poor survival for this disease is often attributable to metastasis[441]. A subset of patients with liver metastases can be treated by surgical resection; nevertheless, even in this subset the majority experience tumour recurrence[442]. Our experiments showing that citrullination of the ECM and expression of PAD4 facilitates human colorectal cancer hepatic metastases may create opportunities for both development of biomarkers and for therapeutic targeting.

In experimental liver metastasis models, we show that genetic downregulation or pharmacological inhibition of PAD4 reduces citrullination and liver metastatic growth. Genetic knockdown of PAD4 prevented formation of liver metastases *in vivo* in the intrasplenic injection model. Notably, downregulation of PAD4 in the cancer cells themselves is sufficient to abrogate liver metastasis without preventing their growth in tissue culture or as a subcutaneous or intercaecal tumour. Pharmacological inhibition of PADs using BB-Cl-amidine had no effect in HCT116 subcutaneous xenografts whereas shRNA knockdown of PAD4 in the same cell line had a substantial effect. This discrepancy could be explained by insufficient bioavailability of BB-Cl-amidine *in vivo* and possible dysregulation of essential cell housekeeping processes in cancer cells depleted of the *PAD4* gene expression. The citrullination identified by proteomics of ECM molecules showed collagens including type I as a target for modification both in

human and in experimental murine liver metastases. Engagement of cancer cells with citrullinated collagen type I altered the cancer cell phenotype with enhanced attachment, decreased motility and gene expression patterns consistent with MET. Whether citrullination of other ECM components has functional consequences remains to be determined by future studies. Importantly, the alterations in gene expression toward increased epithelial markers are induced by plating on citrullinated collagen type I regardless of the intracellular PAD4. These experiments suggest that in colorectal cancer the citrullination of ECM molecules is an important event in altering cancer cell signaling during metastatic growth. In contrast, others have found that reduction in intracellular PAD4 resulted in decreased mesenchymal markers or promoted MET in breast[443] and lung[444] cancer cells. Thus, signaling to affecting EMT markers may be mediated both intra- and extracellularly, reinforcing the possibility that PAD4 and citrullination might be a therapeutic target in liver metastases.

Although our work has focused upon PAD4, we recognise that citrullination may be catalysed by other homologous PAD enzymes, including PADs 1-3 and 6. PAD2 and 4 are most broadly expressed in adults[445] and only PAD4 has a canonical nuclear localisation sequence[446]. Of the PADs, however, PAD4 has been most suggested to be have the potential for involvement in cancer as it is overexpressed in tissues from many different malignancies including colorectal cancer[447], whilst PAD2 expression is downregulated in colorectal cancer compared to normal colon[448]. The release of PAD4 from the cancer cells is clearly complex. Part of this release may be through vesicles such as exosomes. This could be analogous to the release of IL-1β which also lacks a secretory peptide and can be secreted in exosomes, which then through unknown mechanisms release their contents into the extracellular space[437,449,450]. The release through exosomes

is supported by our finding of enrichment of exosomal contents in general within the cancer ECM.

Other molecules we identified in the ECM of metastatic liver also have the capacity to affect EMT. Versican, which was considerably overexpressed in liver metastasis ECM, has been reported to induce MET[451]. TGF-β signaling has been recognised as critical to colorectal cancer liver metastasis[354] and we found upregulation of the TGF-β binding proteins LTBP1-3, TSP1 and TSP2, which contribute to TGF-β activation[452,453]. Further, TGF-β has been linked to tumour suppression through EMT, and its citrullination was reported to alter its activity in the extracellular microenvironment[357,453]. The complex balance between these signals and the function of epithelial-mesenchymal plasticity in cancer is still not completely understood. Genetic manipulation of MET in murine cancer models leads to increased invasion and migration by metastatic cells[454] whilst more recent work implicates MET in promotion of tumour-initiating cells and resistance to chemotherapy[455,456]. After dissemination, MET has been suggested to be necessary to initiate metastatic colony formation[457–459]. Furthermore, MET has been documented in human colorectal cancer liver metastases as demonstrated by Hur et al.[460], showing an increased protein and miR markers for MET in liver metastases compared to their matched primary lesions. These observations are consistent with our *in vivo* results. We further propose consideration of the possibility that cancer cell secretion of PAD4 leads to MET due to modification of the metastatic liver ECM by citrullination with the end result of facilitating metastatic growth. A switch to MET at the distant site in the liver is consistent with our observations that PAD4 expression and increased citrullination are not features of primary colorectal carcinoma.

In conclusion, results presented in this chapter point to MET being enforced by an extracellular signal, that of citrullination of ECM components. This novel mechanism could be exploited by cancer cells to secure their growth at distant sites, and interruption of this signaling could generate an unfavourable microenvironment for cancer cell growth. Therefore, PAD4 and citrullination might be a promising target in therapy of liver metastases.

CHAPTER 5. CONCLUDING REMARKS

5.1. Synopsis

In this thesis, mass spectrometry was utilised to investigate metastatic ECM in terms of evaluating protein composition (chapter 3) and assessing posttranslational modification (chapter 4).

The first aim of chapter 3 was to generate a method to efficiently extract ECM proteins for proteomics analysis. In this work we partially relied on results from my master thesis[250], where I previously optimised a method for mouse liver decellularisation, as well as findings from the pioneering work from Hynes group[251] which utilised compartmental fractionation method to enrich for the ECM. Combination of decellularisation and compartmental fractionation followed by biochemical depletion of nucleic acids and sugars enabled an excellent yield of ECM proteins as documented by silver staining of PAGE-resolved samples. Our improved method therefore allowed a high quality quantitative mass spectrometry analysis of metastatic and unaffected liver tissues.

The second aim of chapter 3 was to compare the abundance of ECM proteins in liver metastasis and uninvolved normal liver using mass spectrometry analysis. Here we first performed a basic qualitative search to obtain a general picture of ECM composition in our samples. In total, we identified 158 matrisome proteins and 939 non-matrix molecules in liver metastasis ECM; however, reclassification of these 939 proteins using GO categories identified many of them as actual ECM- or exosome-related components. We further performed a quantitative analysis of resultant data in order to identify differentially expressed proteins and generated a list of over- and underexpressed ECM constituents of liver metastasis. This chart (Figure 10) mostly contained glycoproteins, among which there were many components known to bind to or interact with TGF-β. Importantly, four proteins exhibited more than a 100 fold change increase as compared to the intact liver – LTBP2, LTBP3, LAMB3 and COL11A1, and each of these proteins was previously linked to cancer progression and metastasis[461–464]. The results obtained by proteomics were further validated using immunoblotting and immunohistochemistry in a separate set of tumour specimens. This testifies to the robustness of our approach and provides a tool which could be utilised by future ECM studies.

The third and final aim of chapter 3 was to use bioinformatics tools for analysis of survival of patients with primary colorectal cancer on the basis of proteomics data. To do so, we examined pathways significantly overrepresented in liver metastasis ECM and then identified that these pathways individually contribute to a worsened survival in colon cancer patients (based on TCGA mRNA dataset). A combined set of 13 genes common for these three pathways significantly predicted disease-free survival in 3 separate bioinformatics datasets (290, 177 and 640 patients). The resultant 13 gene signature can therefore be proposed as a prognostication tool for colorectal adenocarcinoma patients.

Finally, we revealed that the 13 gene signature production is driven by cancer associated fibroblasts and hepatic stellate cells via mTOR and possibly STAT3 pathways in response to factors secreted by cancer cells. Overall, chapter 3 offered a comprehensive proteomics-aided investigation of ECM composition from colorectal cancer liver metastasis and identified a set of ECM genes relevant for progression of this disease.

Having evaluated the composition of the metastatic ECM, we then wished to better characterise posttranslational modifications of the given tumours. The first aim of chapter 4 was to examine if human colorectal liver metastases show increased levels of protein citrullination compared to unaffected tissues. Increased citrullination level in metastasis tissues was independently confirmed using proteomics, immunoblotting and ELISA, all of which demonstrated an increase in citrullinated proteins across tumour samples. In keeping with these findings, citrullination-catalysing enzyme PAD4 was overexpressed in liver metastases, suggesting it could be responsible for inducing higher levels of citrullination. We identified *in vivo* that metastasising cancer cells produce the majority of PAD4 in the tumour microenvironment, and it is delivered to the extracellular space at least in part by means of exosomal transport. Thus, citrullination of the ECM is a characteristic of liver metastasis.

The second and final aim of chapter 4 was to provide an insight on the functional role of citrullination in liver metastasis. Summarising data from literature, this role could be attributed to either immune response or physico-mechanical characteristics. We showed *in vivo* that neutrophil depletion does not affect citrullination in experimental liver metastases, suggesting that this phenomenon is unrelated to NETosis. Seeding cancer cells on citrullinated collagen I altered their adhesive and migratory properties

towards more epithelial phenotype via Integrin- and FAK-mediated mechanism. Inhibition of PAD4 activity *in vivo* reduced intratumoural citrullination as well metastasis burden and epithelial marker expression. Genetic knockdown of PAD4 prevented liver metastases formation in an intrasplenic injection model and reduced metastatic burden in a spontaneous metastasis model. Taken together, these data implicated citrullination/PAD4 in mediating epithelial/mesenchymal plasticity and inhibiting metastatic growth *in vivo*. To summarise, chapter 4 identifies ECM protein citrullination as an important event in colorectal liver metastasis which is required for MET and growth of colonies, raising the possibility of therapeutic targeting of this pathway.

5.2. Study limitations

The results presented in chapters 3 and 4 provide a number of novel findings. Nonetheless, several limitations of this DPhil thesis should be considered. First of all, in spite of successful enrichment of the ECM in chapter 3, it is difficult to estimate or predict the potential loss of protein during the multi-step sample preparation. We detected a considerable variation in the amount of proteins loosely bound to the ECM (e.g. secreted factors), pointing to potential washout of certain classes of molecules. This could result in omission of some proteins and affect the overall reliability and accuracy of the proteomics results. However, we attempted to address this issue by independently validating the observed differences in ECM protein expression using other methods of molecular biology. To further address this problem, alternative and potentially more effective methods of decellularisation could be used; for example, the *ex vivo* portal vein liver perfusion rather than immersion into the decellularisation buffer.

Second, in this study the quantitative proteomics directly compared protein composition of the colorectal liver metastases and uninvolved normal liver, which is not an entirely fair comparison. It is clear that cells comprising the colon and liver are phenotypically different and therefore the observed differences in the ECM composition were somewhat expected and unsurprising. A better approach would be to perform a proteomics-aided comparison of synchronous lesions of colon adenocarcinoma and hepatic metastases resected from the same patient; however, due to technical reasons concerning the difficulty of surgical practice for this disease, sample collection and biobanking, such specimens were virtually impossible to acquire. Interestingly, such analysis was recently performed by Hynes group on a sample of 3 synchronous lesions[465]; however, it was a purely qualitative proteomics study where samples were pooled together thereby reducing the utility of these results.

Third, the overexpression of the 13 gene signature (generated based on the proteomics results) for prediction of survival was based on the mRNA data only; it is well known that mRNA and protein levels do not always correlate in complex biological systems[466], therefore it is difficult to speculate whether the subset of "signature- high" patients truly overexpressed all of the 13 genes. At the moment, there are no publicly available datasets using protein expression for prediction of cancer survival, therefore the above-mentioned shortcoming cannot be addressed for now. Another drawback is that our 13 gene signature was created on the background of the liver metastasis ECM; however, in our study it predicted survival from primary colon adenocarcinoma. It is difficult to draw parallels between primary and metastatic sites of the tumour, since their genetic and phenotypic profiles could often vary significantly[467,468]. In this project we have not examined the expression of the 13 gene signature in primary colorectal

adenocarcinoma lesions and extrapolation of metastasis-derived data to the primary lesions could be counterintuitive or even erroneous. Unfortunately, this issue cannot be addressed at the moment, as none of the publicly available resources enable the survival analysis in a subset of patients with liver metastases nor do they contain databases of colorectal cancer liver metastases. Finally, it is important to note that the observed impact of the 13 gene signature on survival is only correlative and from its analysis we are unable to provide a mechanistic link between the expression of ECM genes and particular biological processes.

Fourth, the methods of studying citrullination both *in vitro* and *in vivo* used in this study are technically challenging and required optimisation. Because of the tight schedule of this project, there was no possibility to perform a more thorough investigation in confirming different hypotheses, involving a wider spectrum of methods and techniques.

Fifth, it is also important to discuss limitations of animal models used in this study. The delivery of tumour cells to the liver by means of intrasplenic injection has several shortcomings. For example, the splenectomy performed after the injection of cells could affect the host immune system, as the spleen is a major depot of neutrophils and monocytes[469]. These changes in the immune system could eventually have an effect on tumour development. However, the splenectomy is absolutely required for this procedure in order to avoid the development of unwanted tumours in the splenic bed. In addition, the injection of a large mass of tumour cells is not physiological and does not recapitulate the natural metastatic dissemination and seeding of cancer cells; moreover, it could create unwanted hemodynamic or inflammatory responses within the liver, which would in turn make this model even less relevant. This model also

does not take into account the systemic effects that a primary tumour has, and which can influence metastasis development. This unfortunately also could not be taken into account by our model. Subcutaneous injection of cancer cells used in this study can also be considered as physiologically different from the process of tumour seeding and growth in humans. Engrafting of tumour cells under the skin has been criticised a lot for unavailability to recapitulate human disease[470], therefore we will not be discussing the shortcomings of this model in this chapter. Finally, intercaecal injection of colon cancer cells used in this study is technically demanding and produces metastases only with few cell lines, including HCT116; there was no possibility to repeat the results using different cell lines due to this reason.

5.3. Future directions

The results presented in this work include data derived from clinical samples, animal studies as well as cell culture experiments. Overall, our findings indicate the importance of the ECM component as well as certain post-translational modifications of the ECM (e.g. citrullination) in the pathogenesis of colorectal cancer and associated liver metastases. Therefore, it is important to consider the potential clinical significance of this work and outline future directions for research.

We identified the ECM signature of 13 genes which predicted poor prognosis in colorectal adenocarcinoma. Although at the moment we cannot speculate on the mechanism underlying this correlation, these data provide an impetus to further investigate cancer ECM for better understanding of tumour biology and development of new prognostication tools. We therefore, on the basis of our findings, suggest the combination of the above-mentioned 13 genes as a potential biomarker for colorectal

carcinoma prognostication after validation of these results using other methods (for example, immunostaining of IHC arrays with available survival and clinical data). Furthermore, in chapter 4 we showed that hepatic liver metastases are characterised by high levels of ECM citrullination. We therefore hypothesise that citrullination might be utilised as a diagnostic biomarker for liver metastases. Currently, anti-citrullinated protein/peptide antibodies (ACPAs, CCP2) are frequently used for diagnosis of RA[471], and it would be of interest to examine whether these commercially available diagnostic antibodies would be successful in detecting hepatic metastases in prospective or retrospective cancer cohorts (although we did not find exactly the same proteins to be citrullinated in liver metastases as in RA). In addition, our studies showed that inhibition of PAD4-driven citrullination using pharmacological inhibitors or genetic knockout delayed growth of liver metastases and reduced metastatic seeding; these results therefore suggest that it would be not unreasonable to pursue further studies to ask whether PAD4/citrullination axis may be targeted for the treatment of hepatic metastases.

REFERENCES

1. Ferlay, J., Parkin, D. M. & Steliarova-Foucher, E. Estimates of cancer incidence and mortality in Europe in 2008. *Eur. J. Cancer* **46,** 765–781 (2010).
2. Jemal, A. *et al.* Global cancer statistics. *CA. Cancer J. Clin.* **61,** 69–90 (2011).
3. Marley, A. R. & Nan, H. Epidemiology of colorectal cancer. *Int. J. Mol. Epidemiol. Genet.* **7,** 105–114 (2016).
4. Theodoratou, E. *et al.* Associations between dietary and lifestyle risk factors and colorectal cancer in the Scottish population. *Eur. J. Cancer Prev.* **23,** 8–17 (2014).
5. Aune, D. *et al.* Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. *BMJ* **343,** d6617 (2011).
6. Aune, D. *et al.* Red and processed meat intake and risk of colorectal adenomas: a systematic review and meta-analysis of epidemiological studies. *Cancer Causes Control* **24,** 611–627 (2013).
7. Johnson, C. M. *et al.* Meta-analyses of colorectal cancer risk factors. *Cancer*

Causes Control **24,** 1207–1222 (2013).

8. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl. Acad. Sci.* **105,** 4283–4288 (2008).

9. Doubeni, C. A. *et al.* Socioeconomic status and the risk of colorectal cancer: an analysis of more than a half million adults in the National Institutes of Health-AARP Diet and Health Study. *Cancer* **118,** 3636–44 (2012).

10. Von Wagner, C. *et al.* Inequalities in participation in an organized national colorectal cancer screening programme: results from the first 2.6 million invitations in England. *Int. J. Epidemiol.* **40,** 712–718 (2011).

11. Rustgi, A. K. The genetics of hereditary colon cancer. *Genes Dev.* **21,** 2525–2538 (2007).

12. Jasperson, K. W., Tuohy, T. M., Neklason, D. W. & Burt, R. W. Hereditary and familial colon cancer. *Gastroenterology* **138,** 2044–58 (2010).

13. Grady, W. M. Genetic testing for high-risk colon cancer patients. *Gastroenterology* **124,** 1574–94 (2003).

14. Mates, I. N. *et al.* Single nucleotide polymorphisms in colorectal cancer: associations with tumor site and TNM stage. *J. Gastrointestin. Liver Dis.* **21,** 45–52 (2012).

15. Ni, J., Wu, G. D., Albenberg, L. & Tomov, V. T. Gut microbiota and IBD: causation or correlation? *Nat. Rev. Gastroenterol. Hepatol.* **14,** 573 (2017).

16. Distrutti, E., Monaldi, L., Ricci, P. & Fiorucci, S. Gut microbiota role in irritable bowel syndrome: New therapeutic strategies. *World J. Gastroenterol.* **22,** 2219–41 (2016).

17. Marasco, G. *et al.* Gut Microbiota and Celiac Disease. *Dig. Dis. Sci.* **61,** 1461–1472 (2016).

18. Kang, M. & Martin, A. Microbiome and colorectal cancer: Unraveling host-microbiota interactions in colitis-associated colorectal cancer development. *Semin. Immunol.* **32,** 3–13 (2017).

19. O'Hara, A. M. & Shanahan, F. The gut flora as a forgotten organ. *EMBO Rep.* **7,** 688–93 (2006).

20. Gagnière, J. *et al.* Gut microbiota imbalance and colorectal cancer. *World J. Gastroenterol.* **22,** 501 (2016).

21. Walia, S., Kamal, R., Dhawan, D. K. & Kanwar, S. S. Chemoprevention by Probiotics During 1,2-Dimethylhydrazine-Induced Colon Carcinogenesis in Rats. *Dig. Dis. Sci.* **63,** 900–909 (2018).

22. Agah, S. *et al.* More Protection of Lactobacillus acidophilus Than Bifidobacterium bifidum Probiotics on Azoxymethane-Induced Mouse Colon Cancer. *Probiotics Antimicrob. Proteins* (2018). doi:10.1007/s12602-018-9425-8

23. Gopalakrishnan, V. *et al.* Gut microbiome modulates response to anti–PD-1 immunotherapy in melanoma patients. *Science (80-.).* **359,** 97–103 (2018).

24. Dai, L. *et al.* Does colorectal cancer significantly influence the assembly of gut microbial communities? *PeerJ* **5,** e3383 (2017).

25. Melville, A., Sheldon, T. A., Gray, R. & Sowden, A. Management of colorectal cancer. *Qual. Health Care* **7,** 103–8 (1998).

26. DE ROSA, M. *et al.* Genetics, diagnosis and management of colorectal cancer (Review). *Oncol. Rep.* **34,** 1087–1096 (2015).

27. Jones, R. P. *et al.* Colorectal Liver Metastases: A Critical Review of State of the Art. *Liver cancer* **6,** 66–71 (2016).

28. Manfredi, S. *et al.* Epidemiology and Management of Liver Metastases From Colorectal Cancer. *Ann. Surg.* **244,** 254–259 (2006).

29. Hackl, C. *et al.* Treatment of colorectal liver metastases in Germany: a ten-year population-based analysis of 5772 cases of primary colorectal adenocarcinoma. *BMC Cancer* **14,** 810 (2014).

30. Sag, A. A., Selcukbiricik, F. & Mandel, N. M. Evidence-based medical oncology and interventional radiology paradigms for liver-dominant colorectal cancer metastases. *World J. Gastroenterol.* **22,** 3127 (2016).

31. Penna, C. & Nordlinger, B. Surgery of liver metastases from colorectal cancer: new promises. *Br. Med. Bull.* **64,** 127–40 (2002).

32. Kemeny, N. & Fata, F. Arterial, portal, or systemic chemotherapy for patients with hepatic metastasis of colorectal carcinoma. *J. Hepatobiliary. Pancreat. Surg.* **6,** 39–49 (1999).

33. van Vledder, M. G. *et al.* Disappearing Colorectal Liver Metastases after Chemotherapy: Should we be Concerned? *J. Gastrointest. Surg.* **14,** 1691–1700 (2010).

34. Benoist, S. *et al.* Complete Response of Colorectal Liver Metastases After Chemotherapy: Does It Mean Cure? *J. Clin. Oncol.* **24,** 3939–3945 (2006).

35. Nordlinger, B. & Rougier, P. Liver metastases from colorectal cancer: the turning point. *J. Clin. Oncol.* **20,** 1442–5 (2002).

36. Leslie, A. & C Steele, R. J. Management of colorectal cancer. *Postgr. Med J* **78,** 473–478 (2002).

37. Diaz-Nieto, R., Fenwick, S., Malik, H. & Poston, G. Defining the Optimal Use of Ablation for Metastatic Colorectal Cancer to the Liver Without High-Level Evidence. *Curr. Treat. Options Oncol.* **18,** 8 (2017).

38. Si-Tayeb, K., Lemaigre, F. P. & Duncan, S. A. Organogenesis and Development of the Liver. *Developmental Cell* (2010). doi:10.1016/j.devcel.2010.01.011

39. Juza, R. M. & Pauli, E. M. Clinical and surgical anatomy of the liver: A review for clinicians. *Clinical Anatomy* (2014). doi:10.1002/ca.22350

40. Malarkey, D. E., Johnson, K., Ryan, L., Boorman, G. & Maronpot, R. R. New Insights into Functional Aspects of Liver Morphology. *Toxicol. Pathol.* (2005). doi:10.1080/01926230590881826

41. Dixon, L. J., Barnes, M., Tang, H., Pritchard, M. T. & Nagy, L. E. in *Comprehensive Physiology* **3,** 785–97 (John Wiley & Sons, Inc., 2013).

42. Puche, J. E., Saiman, Y. & Friedman, S. L. Hepatic stellate cells and liver fibrosis. *Compr. Physiol.* (2013). doi:10.1002/cphy.c120035

43. Tabibian, J. H., Masyuk, A. I., Masyuk, T. V, O'Hara, S. P. & LaRusso, N. F. Physiology of cholangiocytes. *Compr. Physiol.* (2013). doi:10.1002/cphy.c120019

44. Yi, C. X., la Fleur, S. E., Fliers, E. & Kalsbeek, A. The role of the autonomic nervous liver innervation in the control of energy metabolism. *Biochimica et Biophysica Acta - Molecular Basis of Disease* (2010). doi:10.1016/j.bbadis.2010.01.006

45. Jensen, K. J., Alpini, G. & Glaser, S. Hepatic nervous system and neurobiology of the liver. *Compr. Physiol.* (2013). doi:10.1002/cphy.c120018

46. De Feo, P. & Lucidi, P. Liver protein synthesis in physiology and in disease states. *Curr. Opin. Clin. Nutr. Metab. Care* (2002). doi:10.1097/00075197-200201000-00009

47. Ponziani, F. R., Pecere, S., Gasbarrini, A. & Ojetti, V. Physiology and pathophysiology of liver lipid metabolism. *Expert Rev. Gastroenterol. Hepatol.* (2015). doi:10.1586/17474124.2015.1056156

48. Rui, L. Energy metabolism in the liver. *Compr. Physiol.* (2014). doi:10.1002/cphy.c130024

49. Omiecinski, C. J., Vanden Heuvel, J. P., Perdew, G. H. & Peters, J. M. Xenobiotic metabolism, disposition, and regulation by receptors: From biochemical phenomenon to predictors of major toxicities. *Toxicological Sciences* (2011). doi:10.1093/toxsci/kfq338

50. Pantel, K. & Brakenhoff, R. H. Dissecting the metastatic cascade. *Nat. Rev. Cancer* **4,** 448–456 (2004).

51. Valastyan, S. & Weinberg, R. A. Tumor metastasis: molecular insights and evolving paradigms. *Cell* **147,** 275–92 (2011).

52. Su, Z., Yang, Z., Xu, Y., Chen, Y. & Yu, Q. Apoptosis, autophagy, necroptosis, and cancer metastasis. *Mol. Cancer* **14,** 48 (2015).

53. McAllister, S. S. & Weinberg, R. A. The tumour-induced systemic environment as a critical regulator of cancer progression and metastasis. *Nat. Cell Biol.* **16,** 717–727 (2014).

54. Regmi, S., Fu, A. & Luo, K. Q. High Shear Stresses under Exercise Condition Destroy Circulating Tumor Cells in a Microfluidic System. *Sci. Rep.* **7,** 39975 (2017).

55. Mitchell, M. J. & King, M. R. Computational and Experimental Models of Cancer Cell Response to Fluid Shear Stress. *Front. Oncol.* **3,** 44 (2013).

56. Chiang, S. P. H., Cabrera, R. M. & Segall, J. E. Tumor cell intravasation. *Am. J. Physiol. Cell Physiol.* **311,** C1–C14 (2016).

57. Hedley, B. D. & Chambers, A. F. in *Advances in cancer research* **102,** 67–101 (2009).

58. Giancotti, F. G. Mechanisms governing metastatic dormancy and reactivation. *Cell* **155,** 750–64 (2013).

59. Wells, A., Griffith, L., Wells, J. Z. & Taylor, D. P. The dormancy dilemma: quiescence versus balanced proliferation. *Cancer Res.* **73,** 3811–6 (2013).

60. Obenauf, A. C. & Massagué, J. Surviving at a Distance: Organ-Specific Metastasis. *Trends in cancer* **1,** 76–91 (2015).

61. Okuda, R. *et al.* Tropism of cancer stem cells to a specific distant organ. *In Vivo* **28,** 361–5

62. Ring, B. Z. & Ross, D. T. Predicting the sites of metastases. *Genome Biol.* **6,** 241 (2005).

63. Gordon-Weeks, A. N. *et al.* Neutrophils promote hepatic metastasis growth through fibroblast growth factor 2-dependent angiogenesis in mice. *Hepatology* **65,** 1920–1935 (2017).

64. Weidle, U. H., Birzele, F., Kollmorgen, G. & Rüger, R. Molecular Basis of Lung Tropism of Metastasis. *Cancer Genomics Proteomics* **13,** 129–39 (2016).

65. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. in *Nature medicine* **19,** 1423–1437 (2013).

66. Frost, A. R., Hurst, D. R., Shevde, L. A. & Samant, R. S. The Influence of the Cancer Microenvironment on the Process of Metastasis. *Int. J. Breast Cancer* **2012,** 1–3 (2012).

67. Kitamura, T., Qian, B.-Z. & Pollard, J. W. Immune cell promotion of metastasis. *Nat. Rev. Immunol.* **15,** 73–86 (2015).

68. Karagiannis, G. S. *et al.* Cancer-associated fibroblasts drive the progression of metastasis through both paracrine and mechanical pressure on cancer tissue. *Mol. Cancer Res.* **10,** 1403–18 (2012).

69. Paschos, K. A., Majeed, A. W. & Bird, N. C. Role of Kupffer cells in the outgrowth of colorectal cancer liver metastases. *Hepatol. Res.* **40,** 83–94 (2010).

70. Kang, N., Gores, G. J. & Shah, V. H. Hepatic stellate cells: partners in crime for liver metastases? *Hepatology* **54,** 707–13 (2011).

71. Mandal, A. & Viswanathan, C. Natural killer cells: In health and disease. *Hematol. Oncol. Stem Cell Ther.* **8,** 47–55 (2015).

72. Moserle, L. & Casanovas, O. Anti-angiogenesis and metastasis: a tumour and stromal cell alliance. *J. Intern. Med.* **273,** 128–137 (2013).

73. Bielenberg, D. R. & Zetter, B. R. The Contribution of Angiogenesis to the Process of Metastasis. *Cancer J.* **21,** 267–73 (2015).

74. Ribatti, D., Mangialardi, G. & Vacca, A. Stephen Paget and the 'seed and soil' theory of metastatic dissemination. *Clin. Exp. Med.* **6,** 145–149 (2006).

75. Lautt, W. W. *Hepatic Circulation*. *Hepatic Circulation: Physiology and*

Pathophysiology (Morgan & Claypool Life Sciences, 2009).

76. MacPhee, P. J., Schmidt, E. E. & Groom, A. C. Intermittence of blood flow in liver sinusoids, studied by high-resolution in vivo microscopy. *Am. J. Physiol.* **269,** G692-8 (1995).

77. Kumar, A., Sharma, P. & Sarin, S. K. Hepatic venous pressure gradient measurement: time to learn! *Indian J. Gastroenterol.* **27,** 74–80

78. Tiegs, G. & Lohse, A. W. Immune tolerance: what is unique about the liver. *J. Autoimmun.* **34,** 1–6 (2010).

79. Luzzi, K. J. *et al.* Multistep Nature of Metastatic Inefficiency. *Am. J. Pathol.* **153,** 865–873 (1998).

80. Alon, R. & Ley, K. Cells on the run: shear-regulated integrin activation in leukocyte rolling and arrest on endothelial cells. *Curr. Opin. Cell Biol.* **20,** 525–532 (2008).

81. Nielsen, K., Rolff, H. C., Eefsen, R. L. & Vainer, B. The morphological growth patterns of colorectal liver metastases are prognostic for overall survival. *Mod. Pathol.* **27,** 1641–8 (2014).

82. Tremblay, P.-L., Huot, J. & Auger, F. A. Mechanisms by which E-Selectin Regulates Diapedesis of Colon Cancer Cells under Flow Conditions. *Cancer Res.* **68,** 5167–5176 (2008).

83. Vidal-Vanaclocha, F. in 9–42 (2011). doi:10.1007/978-94-007-0292-9_2

84. Vermeulen, P. B. *et al.* Liver metastases from colorectal adenocarcinomas grow in three patterns with different angiogenesis and desmoplasia. *J. Pathol.* **195,** 336–42 (2001).

85. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21,** 1350–1356 (2015).

86. Calon, A. *et al.* Dependency of colorectal cancer on a TGF-β-driven program in stromal cells for metastasis initiation. *Cancer Cell* **22,** 571–84 (2012).

87. Mouw, J. K., Ou, G. & Weaver, V. M. Extracellular matrix assembly: a multiscale deconstruction. *Nat. Rev. Mol. Cell Biol.* **15,** 771–785 (2014).

88. Frantz, C., Stewart, K. M. & Weaver, V. M. The extracellular matrix at a glance. *J. Cell Sci.* **123,** 4195–4200 (2010).

89. Karsdal, M. A. *et al.* Extracellular matrix remodeling: the common denominator in connective tissue diseases. Possibilities for evaluation and current understanding of the matrix as more than a passive architecture, but a key player in tissue failure. *Assay Drug Dev. Technol.* **11,** 70–92 (2013).

90. Pickup, M. W., Mouw, J. K. & Weaver, V. M. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep.* **15,** 1243–53 (2014).

91. Provenzano, P. P. *et al.* Collagen density promotes mammary tumor initiation and progression. *BMC Med.* **6,** 11 (2008).

92. Reinhard, J., Brösicke, N., Theocharidis, U. & Faissner, A. The extracellular

matrix niche microenvironment of neural and cancer stem cells in the brain. *Int. J. Biochem. Cell Biol.* **81,** 174–183 (2016).

93. Wong, G. S. & Rustgi, A. K. Matricellular proteins: priming the tumour microenvironment for cancer development and metastasis. *Br. J. Cancer* **108,** 755–761 (2013).

94. Pickup, M. W., Mouw, J. K. & Weaver, V. M. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep.* **15,** 1243–1253 (2014).

95. Boyd, N. F., Martin, L. J., Yaffe, M. J. & Minkin, S. Mammographic density and breast cancer risk: current understanding and future prospects. *Breast Cancer Res.* **13,** 223 (2011).

96. El-Serag, H. B. Hepatocellular Carcinoma. *N. Engl. J. Med.* **365,** 1118–1127 (2011).

97. Ramirez, F., Sangiorgi, F. O. & Tsipouras, P. Human collagens: biochemical, molecular and genetic features in normal and diseased states. *Horiz. Biochem. Biophys.* **8,** 341–75 (1986).

98. Scott, J. E. Proteoglycan histochemistry--a valuable tool for connective tissue biochemists. *Coll. Relat. Res.* **5,** 541–75 (1985).

99. Lu, P., Takai, K., Weaver, V. M. & Werb, Z. Extracellular Matrix Degradation and Remodeling in Development and Disease. *Cold Spring Harb. Perspect. Biol.* **3,** a005058–a005058 (2011).

100. Reid, S. E. *et al.* Tumor matrix stiffness promotes metastatic cancer cell interaction with the endothelium. *EMBO J.* **36,** 2373–2389 (2017).

101. Schrader, J. *et al.* Matrix stiffness modulates proliferation, chemotherapeutic response, and dormancy in hepatocellular carcinoma cells. *Hepatology* **53,** 1192–205 (2011).

102. Miyamoto, H. *et al.* Tumor-stroma interaction of human pancreatic cancer: acquired resistance to anticancer drugs and proliferation regulation is dependent on extracellular matrix proteins. *Pancreas* **28,** 38–44 (2004).

103. Eke, I., Storch, K., Krause, M. & Cordes, N. Cetuximab Attenuates Its Cytotoxic and Radiosensitizing Potential by Inducing Fibronectin Biosynthesis. *Cancer Res.* **73,** 5869–5879 (2013).

104. Tian, X. *et al.* High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Nature* **499,** 346–349 (2013).

105. Fang, M., Yuan, J., Peng, C. & Li, Y. Collagen as a double-edged sword in tumor progression. *Tumor Biol.* **35,** 2871–2882 (2014).

106. Löffek, S., Franzke, C.-W. & Helfrich, I. Tension in Cancer. *Int. J. Mol. Sci.* **17,** 1910 (2016).

107. Kai, F., Laklai, H. & Weaver, V. M. Force Matters: Biomechanical Regulation of Cell Invasion and Migration in Disease. *Trends Cell Biol.* **26,** 486–497 (2016).

108. Pozzi, A., Yurchenco, P. D. & Iozzo, R. V. The nature and biology of basement membranes. *Matrix Biol.* (2016). doi:10.1016/j.matbio.2016.12.009

109. Shields, M. A., Dangi-Garimella, S., Redig, A. J. & Munshi, H. G. Biochemical role of the collagen-rich tumour microenvironment in pancreatic cancer progression. *Biochem. J.* **441,** 541–552 (2012).

110. Erler, J. T. & Weaver, V. M. Three-dimensional context regulation of metastasis. *Clin. Exp. Metastasis* **26,** 35–49 (2009).

111. Erler, J. T. *et al.* Lysyl oxidase is essential for hypoxia-induced metastasis. *Nature* **440,** 1222–1226 (2006).

112. Levental, K. R. *et al.* Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* **139,** 891–906 (2009).

113. Jiang, H. *et al.* Targeting focal adhesion kinase renders pancreatic cancers responsive to checkpoint immunotherapy. *Nat. Med.* **22,** 851–860 (2016).

114. Valiathan, R. R., Marco, M., Leitinger, B., Kleer, C. G. & Fridman, R. Discoidin domain receptor tyrosine kinases: new players in cancer progression. *Cancer Metastasis Rev.* **31,** 295–321 (2012).

115. Hidalgo-Carcedo, C. *et al.* Collective cell migration requires suppression of actomyosin at cell–cell contacts mediated by DDR1 and the cell polarity regulators Par3 and Par6. *Nat. Cell Biol.* **13,** 49–58 (2011).

116. Zhang, K. *et al.* The collagen receptor discoidin domain receptor 2 stabilizes SNAIL1 to facilitate breast cancer metastasis. *Nat. Cell Biol.* **15,** 677–687 (2013).

117. Boghaert, E. *et al.* Host epithelial geometry regulates breast cancer cell invasiveness. *Proc. Natl. Acad. Sci.* **109,** 19632–19637 (2012).

118. Slattery, M. L. *et al.* Matrix Metalloproteinase Genes Are Associated with Breast Cancer Risk and Survival: The Breast Cancer Health Disparities Study. *PLoS One* **8,** e63165 (2013).

119. Noel, A. *et al.* Membrane associated proteases and their inhibitors in tumour angiogenesis. *J. Clin. Pathol.* **57,** 577–84 (2004).

120. Saeidnia, S., Manayi, A. & Abdollahi, M. From in vitro Experiments to in vivo and Clinical Studies; Pros and Cons. *Curr. Drug Discov. Technol.* **12,** 218–24 (2015).

121. Crotti, S. *et al.* Extracellular Matrix and Colorectal Cancer: How Surrounding Microenvironment Affects Cancer Cell Behavior? *J. Cell. Physiol.* **232,** 967–975 (2017).

122. Pocard, M., Tsukui, H., Salmon, R. J., Dutrillaux, B. & Poupon, M. F. Efficiency of orthotopic xenograft models for human colon cancers. *In Vivo* **10,** 463–469 (1996).

123. Tseng, W., Leong, X. & Engleman, E. Orthotopic Mouse Model of Colorectal Cancer. *J. Vis. Exp.* (2007). doi:10.3791/484

124. Lorenz, E. & Stewart, H. L. Intestinal Carcinoma and Other Lesions in Mice Following Oral Administration of 1,2,5,6-Dibenzanthracene and 20-Methylcholanthrene. *JNCI J. Natl. Cancer Inst.* **1,** 17–40 (1940).

125. Walpole, A. L., Williams, M. H. C. & Roberts, D. C. The carcinogenic action of 4-aminodiphenyl and 3:2'-dimethyl-4-amino-diphenyl. *Br. J. Ind. Med.* **9,** 255–263 (1952).

126. Reddy, B. S. & Ohmori, T. Effect of intestinal microflora and dietary fat on 3,2'-dimethyl-4-aminobiphenyl-induced colon carcinogenesis in F344 rats. *Cancer Res.* **41,** 1363–1367 (1981).

127. Jägerstad, M. *et al.* Effects of meat composition and cooking conditions on the formation of mutagenic imidazoquinoxalines (MeIQx and its methyl derivatives). *Princess Takamatsu Symp.* **16,** 87–96 (1985).

128. Sugimura, T., Wakabayashi, K., Nakagama, H. & Nagao, M. Heterocyclic amines: Mutagens/carcinogens produced during cooking of meat and fish. *Cancer Sci.* **95,** 290–299 (2004).

129. Hasegawa, R. *et al.* Dose-dependence of 2-amino-1-methyl-6-phenylimidazo[4,5-b]-pyridine (PhIP) carcinogenicity in rats. *Carcinogenesis* **14,** 2553–2557 (1993).

130. Ito, N. *et al.* A new colon and mammary carcinogen in cooked food, 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP). *Carcinogenesis* **12,** 1503–1506 (1991).

131. Ghoshal, A., Preisegger, K. H., Takayama, S., Thorgeirsson, S. S. & Snyderwine, E. G. Induction of mammary tumors in female Sprague-Dawley rats by the food-derived carcinogen 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine and effect of dietary fat. *Carcinogenesis* **15,** 2429–2433 (1994).

132. Kobaek-Larsen, M., Thorup, I., Diederichsen, A., Fenger, C. & Hoitinga, M. R. Review of colorectal cancer and its metastases in rodent models: comparative aspects with those in humans. *Comp. Med.* **50,** 16–26 (2000).

133. Reddy, B. S. & Rivenson, A. Inhibitory effect of Bifidobacterium longum on colon, mammary, and liver carcinogenesis induced by 2-amino-3-methylimidazo[4,5-f]quinoline, a food mutagen. *Cancer Res.* **53,** 3914–3918 (1993).

134. Ohgaki, H. *et al.* Carcinogenicity in mice and rats of heterocyclic amines in cooked foods. *Environ. Health Perspect.* **67,** 129–134 (1986).

135. Laqueur, G. L., Mickelsen, O., Whiting, M. G. & Kurland, L. T. CARCINOGENIC PROPERTIES OF NUTS FROM CYCAS CIRCINALIS L. INDIGENOUS TO GUAM. *J. Natl. Cancer Inst.* **31,** 919–951 (1963).

136. Laqueur, G. L. CARCINOGENIC EFFECTS OF CYCAD MEAL AND CYCASIN, METHYLAZOXYMETHANOL GLYCOSIDE, IN RATS AND EFFECTS OF CYCASIN IN GERMFREE RATS. *Fed. Proc.* **23,** 1386–1388 (1964).

137. Fiala, E. S. Investigations into the metabolism and mode of action of the colon carcinogens 1,2-dimethylhydrazine and azoxymethane. *Cancer* **40,** 2436–2445 (1977).

138. Fiala, E. S., Kulakis, C., Christiansen, G. & Weisburger, J. H. Inhibition of the metabolism of the colon carcinogen, azoxymethane, by pyrazole. *Cancer Res.*

38, 4515–4521 (1978).

139. Femia, A. Pietro *et al.* Mucin-depleted foci have beta-catenin gene mutations, altered expression of its protein, and are dose- and time-dependent in the colon of 1,2-dimethylhydrazine-treated rats. *Int. J. Cancer* **116,** 9–15 (2005).

140. Bolt, A. B., Papanikolaou, A., Delker, D. A., Wang, Q. S. & Rosenberg, D. W. Azoxymethane induces KI-ras activation in the tumor resistant AKR/J mouse colon. *Mol. Carcinog.* **27,** 210–218 (2000).

141. Jacoby, R. F., Llor, X., Teng, B. B., Davidson, N. O. & Brasitus, T. A. Mutations in the K-ras oncogene induced by 1,2-dimethylhydrazine in preneoplastic and neoplastic rat colonic mucosa. *J. Clin. Invest.* **87,** 624–630 (1991).

142. Takahashi, M., Mutoh, M., Kawamori, T., Sugimura, T. & Wakabayashi, K. Altered expression of beta-catenin, inducible nitric oxide synthase and cyclooxygenase-2 in azoxymethane-induced rat colon carcinogenesis. *Carcinogenesis* **21,** 1319–1327 (2000).

143. Maltzman, T., Whittington, J., Driggers, L., Stephens, J. & Ahnen, D. AOM-induced mouse colon tumors do not express full-length APC protein. *Carcinogenesis* **18,** 2435–2439 (1997).

144. Guda, K., Giardina, C., Nambiar, P., Cui, H. & Rosenberg, D. W. Aberrant transforming growth factor-beta signaling in azoxymethane-induced mouse colon tumors. *Mol. Carcinog.* **31,** 204–213 (2001).

145. Clevers, H. Wnt/beta-catenin signaling in development and disease. *Cell* **127,** 469–480 (2006).

146. Moser, A. R., Pitot, H. C. & Dove, W. F. A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. *Science* **247,** 322–324 (1990).

147. McCart, A. E., Vickaryous, N. K. & Silver, A. Apc mice: models, modifiers and mutants. *Pathol. Res. Pract.* **204,** 479–490 (2008).

148. Romagnolo, B. *et al.* Intestinal dysplasia and adenoma in transgenic mice after overexpression of an activated beta-catenin. *Cancer Res.* **59,** 3875–3879 (1999).

149. Harada, N. *et al.* Intestinal polyposis in mice with a dominant stable mutation of the beta-catenin gene. *EMBO J.* **18,** 5931–5942 (1999).

150. Forrester, K., Almoguera, C., Han, K., Grizzle, W. E. & Perucho, M. Detection of high incidence of K-ras oncogenes during human colon tumorigenesis. *Nature* **327,** 298–303 (1987).

151. Johnson, L. *et al.* Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. *Nature* **410,** 1111–1116 (2001).

152. Janssen, K.-P. *et al.* Targeted expression of oncogenic K-ras in intestinal epithelium causes spontaneous tumorigenesis in mice. *Gastroenterology* **123,** 492–504 (2002).

153. Lampropoulos, P. *et al.* TGF-beta signalling in colon carcinogenesis. *Cancer Lett.* **314,** 1–7 (2012).

154. Zhu, Y., Richardson, J. A., Parada, L. F. & Graff, J. M. Smad3 mutant mice

develop metastatic colorectal cancer. *Cell* **94,** 703–714 (1998).

155. Maggio-Price, L. *et al.* Helicobacter infection is required for inflammation and colon cancer in SMAD3-deficient mice. *Cancer Res.* **66,** 828–838 (2006).

156. Wang, T.-H., Hsia, S.-M. & Shieh, T.-M. Lysyl Oxidase and the Tumor Microenvironment. *Int. J. Mol. Sci.* **18,** 62 (2016).

157. Zhang, B. *et al.* Tumor-derived matrix metalloproteinase-13 (MMP-13) correlates with poor prognosis of invasive breast cancer. *BMC Cancer* **8,** 83 (2008).

158. Têtu, B. *et al.* The influence of MMP-14, TIMP-2 and MMP-2 expression on breast cancer prognosis. *Breast Cancer Res.* **8,** R28 (2006).

159. Freire-de-Lima, L. *et al.* Involvement of O-glycosylation defining oncofetal fibronectin in epithelial-mesenchymal transition process. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 17690–5 (2011).

160. Richter, P. *et al.* IIICS de novo glycosylated fibronectin as a marker for invasiveness in urothelial carcinoma of the urinary bladder (UBC). *J. Cancer Res. Clin. Oncol.* **134,** 1059–1065 (2008).

161. Leeming, D. J. *et al.* Post-translational modifications of the extracellular matrix are key events in cancer progression: Opportunities for biochemical marker development. *Biomarkers* **16,** 193–205 (2011).

162. Harauz, G. & Musse, A. a. A tale of two citrullines--structural and functional aspects of myelin basic protein deimination in health and disease. *Neurochem. Res.* **32,** 137–58 (2007).

163. Ghari, F. *et al.* Citrullination-acetylation interplay guides E2F-1 activity during the inflammatory response. *Sci. Adv.* **2,** e1501257 (2016).

164. Tessarz, P. & Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* **15,** 703–708 (2014).

165. Vossenaar, E. R., Zendman, A. J. W., Van Venrooij, W. J. & Pruijn, G. J. M. PAD, a growing family of citrullinating enzymes: Genes, features and involvement in disease. *BioEssays* **25,** 1106–1118 (2003).

166. Husson, A., Brasse-Lagnel, C., Fairand, A., Renouf, S. & Lavoinne, A. Argininosuccinate synthetase from the urea cycle to the citrulline-NO cycle. *Eur. J. Biochem.* **270,** 1887–99 (2003).

167. Rogers, G. E., Harding, H. W. & Llewellyn-Smith, I. J. The origin of citrulline-containing proteins in the hair follicle and the chemical nature of trichohyalin, an intracellular precursor. *Biochim. Biophys. Acta* **495,** 159–75 (1977).

168. Knuckley, B., Bhatia, M. & Thompson, P. R. Protein Arginine Deiminase 4: Evidence for a Reverse Protonation Mechanism †. *Biochemistry* **46,** 6578–6587 (2007).

169. Nicholas, A. P. & Bhattacharya, S. K. *Protein deimination in human health and disease*. *Protein Deimination in Human Health and Disease* (2014). doi:10.1007/978-1-4614-8317-5

170. Nguyen, H. & James, E. A. Immune recognition of citrullinated epitopes. *Immunology* **149,** 131–138 (2016).

171. Konig, M. F. *et al.* Aggregatibacter actinomycetemcomitans-induced hypercitrullination links periodontal infection to autoimmunity in rheumatoid arthritis. *Sci. Transl. Med.* **8,** 369ra176-369ra176 (2016).

172. Carrillo-Vico, A., Leech, M. D. & Anderton, S. M. Contribution of Myelin Autoantigen Citrullination to T Cell Autoaggression in the Central Nervous System. *J. Immunol.* **184,** 2839–2846 (2010).

173. Pruijn, G. J. M. Citrullination and Carbamylation in the Pathophysiology of Rheumatoid Arthritis. *Front. Immunol.* **6,** 192 (2015).

174. Klareskog, L. *et al.* A new model for an etiology of rheumatoid arthritis: Smoking may trigger HLA–DR (shared epitope)–restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum.* **54,** 38–46 (2006).

175. Ireland, J., Herzog, J. & Unanue, E. R. Cutting edge: unique T cells that recognize citrullinated peptides are a feature of protein immunization. *J. Immunol.* **177,** 1421–5 (2006).

176. Scally, S. W. *et al.* A molecular basis for the association of the *HLA-DRB1* locus, citrullination, and rheumatoid arthritis. *J. Exp. Med.* **210,** 2569–2582 (2013).

177. Snir, O. *et al.* Identification and functional characterization of T cells reactive to citrullinated vimentin in HLA-DRB1*0401-positive humanized mice and rheumatoid arthritis patients. *Arthritis Rheum.* **63,** 2873–2883 (2011).

178. Freed, B. M., Schuyler, R. P. & Aubrey, M. T. Association of the HLA-DRB1 epitope LA67, 74 with rheumatoid arthritis and citrullinated vimentin binding. *Arthritis Rheum.* **63,** 3733–3739 (2011).

179. Gerstner, C. *et al.* Functional and Structural Characterization of a Novel HLA-DRB1*04:01-Restricted α-Enolase T Cell Epitope in Rheumatoid Arthritis. *Front. Immunol.* **7,** 494 (2016).

180. György, B., Tóth, E., Tarcsa, E., Falus, A. & Buzás, E. I. Citrullination: A posttranslational modification in health and disease. *Int. J. Biochem. Cell Biol.* **38,** 1662–1677 (2006).

181. van Venrooij, W. J. & Zendman, A. J. W. Anti-CCP2 antibodies: an overview and perspective of the diagnostic abilities of this serological marker for early rheumatoid arthritis. *Clin. Rev. Allergy Immunol.* **34,** 36–9 (2008).

182. Aggarwal, R., Liao, K., Nair, R., Ringold, S. & Costenbader, K. H. Anti-citrullinated peptide antibody assays and their role in the diagnosis of rheumatoid arthritis. *Arthritis Rheum.* **61,** 1472–83 (2009).

183. Vossenaar, E. R. *et al.* The presence of citrullinated proteins is not specific for rheumatoid synovial tissue. *Arthritis Rheum.* **50,** 3485–3494 (2004).

184. Kastbom, A., Strandberg, G., Lindroos, A. & Skogh, T. Anti-CCP antibody test predicts the disease course during 3 years in early rheumatoid arthritis (the Swedish TIRA project). *Ann. Rheum. Dis.* **63,** 1085–9 (2004).

185. Turesson, C., Eberhardt, K., Jacobsson, L. T. H. & Lindqvist, E. Incidence and predictors of severe extra-articular disease manifestations in an early rheumatoid arthritis inception cohort. *Ann. Rheum. Dis.* **66,** 1543–1544 (2007).

186. Geraldino-Pardilla, L. *et al.* Association of Anti-Citrullinated Peptide Antibodies with Coronary Artery Calcification in Rheumatoid Arthritis. *Arthritis Care Res. (Hoboken).* (2016). doi:10.1002/acr.23106

187. Vallbracht, I. *et al.* Diagnostic and clinical value of anti-cyclic citrullinated peptide antibodies compared with rheumatoid factor isotypes in rheumatoid arthritis. *Ann. Rheum. Dis.* **63,** 1079–84 (2004).

188. Kuhn, K. A. *et al.* Antibodies against citrullinated proteins enhance tissue injury in experimental autoimmune arthritis. *J. Clin. Invest.* **116,** 961–973 (2006).

189. Kawalkowska, J. *et al.* Abrogation of collagen-induced arthritis by a peptidyl arginine deiminase inhibitor is associated with modulation of T cell-mediated immune responses. *Sci. Rep.* **6,** 26430 (2016).

190. Willis, V. C. *et al.* N-α-benzoyl-N5-(2-chloro-1-iminoethyl)-L-ornithine amide, a protein arginine deiminase inhibitor, reduces the severity of murine collagen-induced arthritis. *J. Immunol.* **186,** 4396–404 (2011).

191. Seri, Y. *et al.* Peptidylarginine deiminase type 4 deficiency reduced arthritis severity in a glucose-6-phosphate isomerase-induced arthritis model. *Sci. Rep.* **5,** 13041 (2015).

192. Suzuki, A. *et al.* Decreased severity of experimental autoimmune arthritis in peptidylarginine deiminase type 4 knockout mice. *BMC Musculoskelet. Disord.* **17,** 205 (2016).

193. Brinkmann, V. *et al.* Neutrophil Extracellular Traps Kill Bacteria. *Science (80-.).* **303,** 1532–1535 (2004).

194. Sur Chowdhury, C. *et al.* Enhanced neutrophil extracellular trap generation in rheumatoid arthritis: analysis of underlying signal transduction pathways and potential diagnostic utility. *Arthritis Res. Ther.* **16,** R122 (2014).

195. Papadaki, G. *et al.* Enhanced release of neutrophil extracellular traps from peripheral blood neutrophils in patients with rheumatoid arthritis. *Ann. Rheum. Dis.* **71,** A79.2-A79 (2012).

196. Zhong, X.-Y. *et al.* Increased Concentrations of Antibody-Bound Circulatory Cell-Free DNA in Rheumatoid Arthritis. *Clin. Chem.* **53,** 1609–1614 (2007).

197. Khandpur, R. *et al.* NETs are a source of citrullinated autoantigens and stimulate inflammatory responses in rheumatoid arthritis. *Sci. Transl. Med.* **5,** 178ra40 (2013).

198. Papadaki, G. *et al.* Neutrophil extracellular traps exacerbate Th1-mediated autoimmune responses in rheumatoid arthritis by promoting DC maturation. *Eur. J. Immunol.* **46,** 2542–2554 (2016).

199. Cools-Lartigue, J. *et al.* Neutrophil extracellular traps sequester circulating tumor cells and promote metastasis. *J. Clin. Invest.* **123,** 3446–3458 (2013).

200. Tohme, S. *et al.* Neutrophil Extracellular Traps Promote the Development and Progression of Liver Metastases after Surgical Stress. *Cancer Res.* **76,** 1367–1380 (2016).

201. Hakkim, A. *et al.* Impairment of neutrophil extracellular trap degradation is associated with lupus nephritis. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 9813–8 (2010).

202. Kaplan, M. J. & Radic, M. Neutrophil Extracellular Traps: Double-Edged Swords of Innate Immunity. *J. Immunol.* **189,** 2689–2695 (2012).

203. Brinkmann, V. & Zychlinsky, A. Neutrophil extracellular traps: Is immunity the second function of chromatin? *J. Cell Biol.* **198,** (2012).

204. Villanueva, E. *et al.* Netting Neutrophils Induce Endothelial Damage, Infiltrate Tissues, and Expose Immunostimulatory Molecules in Systemic Lupus Erythematosus. *J. Immunol.* **187,** 538–552 (2011).

205. Knight, J. S. *et al.* Peptidylarginine deiminase inhibition disrupts NET formation and protects against kidney, skin and vascular disease in lupus-prone MRL/ *lpr* mice. *Ann. Rheum. Dis.* **74,** 2199–2206 (2015).

206. Knight, J. S. *et al.* Peptidylarginine Deiminase Inhibition Reduces Vascular Damage and Modulates Innate Immune Responses in Murine Models of Atherosclerosis. *Circ. Res.* **114,** 947–956 (2014).

207. Ge, L. *et al.* Neutrophil extracellular traps in ischemia-reperfusion injury-induced myocardial no-reflow: therapeutic potential of DNase-based reperfusion strategy. *Am. J. Physiol. - Hear. Circ. Physiol.* **308,** (2015).

208. Chumanevich, A. A. *et al.* Suppression of colitis in mice by Cl-amidine: a novel peptidylarginine deiminase inhibitor. *AJP Gastrointest. Liver Physiol.* **300,** G929–G938 (2011).

209. Wood, D. D., Moscarello, M. A., Bilbao, J. M. & O'Connors, P. Acute multiple sclerosis (marburg type) is associated with developmentally immature myelin basic protein. *Ann. Neurol.* **40,** 18–24 (1996).

210. Nicholas, A. P. *et al.* Immunohistochemical localization of citrullinated proteins in adult rat brain. *J. Comp. Neurol.* **459,** 251–66 (2003).

211. Cao, L. *et al.* Inhibition of experimental allergic encephalomyelitis in the Lewis rat by paclitaxel. *J. Neuroimmunol.* **108,** 103–11 (2000).

212. Beniac, D. R. *et al.* Cryoelectron Microscopy of Protein–Lipid Complexes of Human Myelin Basic Protein Charge Isomers Differing in Degree of Citrullination. *J. Struct. Biol.* **129,** 80–95 (2000).

213. Inagaki, M., Takahara, H., Nishi, Y., Sugawara, K. & Sato, C. Ca2+-dependent deimination-induced disassembly of intermediate filaments involves specific modification of the amino-terminal head domain. *J. Biol. Chem.* **264,** 18119–27 (1989).

214. Mastronardi, F. G. *et al.* Increased Citrullination of Histone H3 in Multiple Sclerosis Brain and Animal Models of Demyelination: A Role for Tumor Necrosis Factor-Induced Peptidylarginine Deiminase 4 Translocation. *J.*

Neurosci. **26,** 11387–11396 (2006).

215. Mastronardi, F. G., Noor, A., Wood, D. D., Paton, T. & Moscarello, M. A. Peptidyl argininedeiminase 2 CpG island in multiple sclerosis white matter is hypomethylated. *J. Neurosci. Res.* **85,** 2006–16 (2007).

216. Musse, A. A. *et al.* Peptidylarginine deiminase 2 (PAD2) overexpression in transgenic mice leads to myelin loss in the central nervous system. *Dis. Model. Mech.* **1,** 229–40 (2008).

217. Jang, B. *et al.* Accumulation of Citrullinated Proteins by Up-Regulated Peptidylarginine Deiminase 2 in Brains of Scrapie-Infected Mice. *Am. J. Pathol.* **173,** 1129–1142 (2008).

218. Jang, B. *et al.* Involvement of peptidylarginine deiminase-mediated post-translational citrullination in pathogenesis of sporadic Creutzfeldt-Jakob disease. *Acta Neuropathol.* **119,** 199–210 (2010).

219. Ishigami, A. *et al.* Abnormal accumulation of citrullinated proteins catalyzed by peptidylarginine deiminase in hippocampal extracts from patients with Alzheimer's disease. *J. Neurosci. Res.* **80,** 120–128 (2005).

220. Lange, S. *et al.* Protein deiminases: new players in the developmentally regulated loss of neural regenerative ability. *Dev. Biol.* **355,** 205–14 (2011).

221. Li, P. *et al.* Regulation of p53 Target Gene Expression by Peptidylarginine Deiminase 4. *Mol. Cell. Biol.* **28,** 4745–4758 (2008).

222. Yao, H. *et al.* Histone Arg Modifications and p53 Regulate the Expression of OKL38, a Mediator of Apoptosis. *J. Biol. Chem.* **283,** 20060–20068 (2008).

223. Li, P. *et al.* Coordination of PAD4 and HDAC2 in the regulation of p53-target gene expression. *Oncogene* **29,** 3153–3162 (2010).

224. Guo, Q. & Fast, W. Citrullination of Inhibitor of Growth 4 (ING4) by Peptidylarginine Deminase 4 (PAD4) Disrupts the Interaction between ING4 and p53. *J. Biol. Chem.* **286,** 17069–17078 (2011).

225. Tanikawa, C. *et al.* Regulation of histone modification and chromatin structure by the p53-PADI4 pathway. *Nat. Commun.* **3,** 676 (2012).

226. Cook, K. *et al.* Citrullinated α-enolase is an effective target for anti-cancer immunity. *Oncoimmunology* **7,** e1390642 (2018).

227. Funayama, R. *et al.* Protein-arginine deiminase 2 suppresses proliferation of colon cancer cells through protein citrullination. *Cancer Sci.* **108,** 713–718 (2017).

228. Chang, X. *et al.* Increased PADI4 expression in blood and tissues of patients with malignant tumors. *BMC Cancer* **9,** 40 (2009).

229. Zhang, X. *et al.* Genome-wide analysis reveals PADI4 cooperates with Elk-1 to activate c-Fos expression in breast cancer cells. *PLoS Genet.* **7,** e1002112 (2011).

230. Zhang, X. *et al.* Genome-Wide Analysis Reveals PADI4 Cooperates with Elk-1 to Activate c-Fos Expression in Breast Cancer Cells. *PLoS Genet.* **7,** e1002112 (2011).

231. Chai, Y. *et al.* c-Fos oncogene regulator Elk-1 interacts with BRCA1 splice variants BRCA1a/1b and enhances BRCA1a/1b-mediated growth suppression in breast cancer cells. *Oncogene* **20,** 1357–1367 (2001).

232. Cantarino, N. *et al.* Downregulation of the Deiminase PADI2 Is an Early Event in Colorectal Carcinogenesis and Indicates Poor Prognosis. *Mol. Cancer Res.* **14,** 841–848 (2016).

233. Cherrington, B. D. *et al.* Comparative analysis of peptidylarginine deiminase-2 expression in canine, feline and human mammary tumours. *J. Comp. Pathol.* **147,** 139–46 (2012).

234. Senshu, T., Akiyama, K. & Nomura, K. Identification of citrulline residues in the V subdomains of keratin K1 derived from the cornified layer of newborn mouse epidermis. *Exp. Dermatol.* **8,** 392–401 (1999).

235. Senshu, T. *et al.* Detection of deiminated proteins in rat skin: probing with a monospecific antibody after modification of citrulline residues. *J. Invest. Dermatol.* **105,** 163–9 (1995).

236. Senshu, T., Akiyama, K., Ishigami, A. & Nomura, K. Studies on specificity of peptidylarginine deiminase reactions using an immunochemical probe that recognizes an enzymatically deiminated partial sequence of mouse keratin K1. *J. Dermatol. Sci.* **21,** 113–26 (1999).

237. Nicholas, A. P. & Whitaker, J. N. Preparation of a monoclonal antibody to citrullinated epitopes: its characterization and some applications to immunohistochemistry in human brain. *Glia* **37,** 328–36 (2002).

238. Bicker, K. L., Subramanian, V., Chumanevich, A. A., Hofseth, L. J. & Thompson, P. R. Seeing Citrulline: Development of a Phenylglyoxal-Based Probe To Visualize Protein Citrullination. *J. Am. Chem. Soc.* **134,** 17015–17018 (2012).

239. Siuzdak, G. The Expanding Role of Mass Spectrometry in Biotechnology. *Expand. role mass Spectrom. Biotechnol.* **15,** 625 (2006).

240. Kubota, K., Yoneyama-Takazawa, T. & Ichikawa, K. Determination of sites citrullinated by peptidylarginine deiminase using 18O stable isotope labeling and mass spectrometry. *Rapid Commun. Mass Spectrom.* **19,** 683–8 (2005).

241. Goeb, V. *et al.* Candidate autoantigens identified by mass spectrometry in early rheumatoid arthritis are chaperones and citrullinated glycolytic enzymes. *Arthritis Res Ther* **11,** R38 (2009).

242. Schwenzer, A. *et al.* Identification of an immunodominant peptide from citrullinated tenascin-C as a major target for autoantibodies in rheumatoid arthritis. *Ann. Rheum. Dis.* **75,** 1876–1883 (2016).

243. Grant, J. E. & Li, H. in *Protein Deimination in Human Health and Disease* 347–365 (Springer New York, 2014). doi:10.1007/978-1-4614-8317-5_19

244. Subramanian, V., Slade, D. J. & Thompson, P. R. in *Protein Deimination in Human Health and Disease* 377–427 (Springer New York, 2014). doi:10.1007/978-1-4614-8317-5_21

245. Kawalkowska, J. *et al.* Abrogation of collagen-induced arthritis by a peptidyl arginine deiminase inhibitor is associated with modulation of T cell-mediated immune responses. *Sci. Rep.* **6,** 26430 (2016).

246. Knight, J. S. *et al.* Peptidylarginine deiminase inhibition disrupts NET formation and protects against kidney, skin and vascular disease in lupus-prone MRL/ *lpr* mice. *Ann. Rheum. Dis.* **74,** 2199–2206 (2015).

247. Pritzker, L. B. & Moscarello, M. A. A novel microtubule independent effect of paclitaxel: the inhibition of peptidylarginine deiminase from bovine brain. *Biochim. Biophys. Acta* **1388,** 154–60 (1998).

248. Hidaka, Y., Hagiwara, T. & Yamada, M. *Methylation of the guanidino group of arginine residues prevents citrullination by peptidylarginine deiminase IV*. *FEBS Letters* **579,** (2005).

249. Workman, P. *et al.* Guidelines for the welfare and use of animals in cancer research. *Br. J. Cancer* **102,** 1555–1577 (2010).

250. Yuzhalin, A. E. Proteomics-aided search for extracellular matrix signatures of colorectal hepatic metastases ex vivo. (2014).

251. Naba, A. *et al.* The Matrisome: *In Silico* Definition and *In Vivo* Characterization by Proteomics of Normal and Tumor Extracellular Matrices. *Mol. Cell. Proteomics* **11,** M111.014647 (2012).

252. Mi, H. *et al.* The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* **33,** D284–D288 (2004).

253. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discov.* **2,** 401–404 (2012).

254. Li, Q., Birkbak, N. J., Gyorffy, B., Szallasi, Z. & Eklund, A. C. Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics* **12,** 474 (2011).

255. Crapo, P. M., Gilbert, T. W. & Badylak, S. F. An overview of tissue and whole organ decellularization processes. *Biomaterials* **32,** 3233–43 (2011).

256. Bonnans, C., Chou, J. & Werb, Z. Remodelling the extracellular matrix in development and disease. *Nat. Rev. Mol. Cell Biol.* **15,** 786–801 (2014).

257. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144,** 646–674 (2011).

258. Ricard-Blum, S. The collagen family. *Cold Spring Harb. Perspect. Biol.* **3,** a004978 (2011).

259. Shoulders, M. D. & Raines, R. T. Collagen Structure and Stability. *Annu. Rev. Biochem.* **78,** 929–958 (2009).

260. Herchenhan, A. *et al.* Lysyl Oxidase Activity Is Required for Ordered Collagen Fibrillogenesis by Tendon Cells. *J. Biol. Chem.* **290,** 16440–50 (2015).

261. Schneider, V. A. & Granato, M. The Myotomal diwanka (lh3) Glycosyltransferase and Type XVIII Collagen Are Critical for Motor Growth

Cone Migration. *Neuron* **50,** 683–695 (2006).

262. Myllyharju, J. Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol.* **22,** 15–24 (2003).

263. Winter, A. D., McCormack, G. & Page, A. P. Protein disulfide isomerase activity is essential for viability and extracellular matrix formation in the nematode Caenorhabditis elegans. *Dev. Biol.* **308,** 449–461 (2007).

264. EASTOE, J. E. The amino acid composition of mammalian collagen and gelatin. *Biochem. J.* **61,** 589–600 (1955).

265. Somaiah, C. *et al.* Collagen Promotes Higher Adhesion, Survival and Proliferation of Mesenchymal Stem Cells. *PLoS One* **10,** e0145068 (2015).

266. Koohestani, F. *et al.* Extracellular matrix collagen alters cell proliferation and cell cycle progression of human uterine leiomyoma smooth muscle cells. *PLoS One* **8,** e75844 (2013).

267. Rasmussen, C. H., Petersen, D. R., Moeller, J. B., Hansson, M. & Dufva, M. Collagen Type I Improves the Differentiation of Human Embryonic Stem Cells towards Definitive Endoderm. *PLoS One* **10,** e0145389 (2015).

268. Yannas, I. V. in 219–244 (Springer, Berlin, Heidelberg, 1995). doi:10.1007/3540587888_17

269. Fassett, J., Tobolt, D. & Hansen, L. K. Type I collagen structure regulates cell morphology and EGF signaling in primary rat hepatocytes through cAMP-dependent protein kinase A. *Mol. Biol. Cell* **17,** 345–56 (2006).

270. Grzesiak, J. J., Pierschbacher, M. D., Amodeo, M. F., Malaney, T. I. & Glass, J. R. Enhancement of cell interactions with collagen/glycosaminoglycan matrices by RGD derivatization. *Biomaterials* **18,** 1625–32 (1997).

271. Canty, E. G. & Kadler, K. E. Procollagen trafficking, processing and fibrillogenesis. *J. Cell Sci.* **118,** 1341–53 (2005).

272. Myllyharju, J. & Kivirikko, K. I. Collagens and collagen-related diseases. *Ann. Med.* **33,** 7–21 (2001).

273. Varani, J. *et al.* Decreased collagen production in chronologically aged skin: roles of age-dependent alteration in fibroblast function and defective mechanical stimulation. *Am. J. Pathol.* **168,** 1861–8 (2006).

274. Singh, O. *et al.* Collagen dressing versus conventional dressings in burn and chronic wounds: a retrospective study. *J. Cutan. Aesthet. Surg.* **4,** 12–6 (2011).

275. Vellinga, T. T. *et al.* Collagen-rich stroma in aggressive colon tumors induces mesenchymal gene expression and tumor cell invasion. *Oncogene* **35,** 5263–5271 (2016).

276. Galván, J. A. *et al.* Validation of COL11A1/procollagen 11A1 expression in TGF-β1-activated immortalised human mesenchymal cells and in stromal cells of human colon adenocarcinoma. *BMC Cancer* **14,** 867 (2014).

277. Moilanen, J. M. *et al.* Collagen XVII expression correlates with the invasion and metastasis of colorectal cancer. *Hum. Pathol.* **46,** 434–442 (2015).

278. Skovbjerg, H. *et al.* Collagen mRNA levels changes during colorectal cancer carcinogenesis. *BMC Cancer* **9,** 136 (2009).

279. Zhang, Z., Wang, Y., Zhang, J., Zhong, J. & Yang, R. COL1A1 promotes metastasis in colorectal cancer by regulating the WNT/PCP pathway. *Mol. Med. Rep.* (2018). doi:10.3892/mmr.2018.8533

280. Blockhuys, S. *et al.* X-radiation enhances the collagen type I strap formation and migration potentials of colon cancer cells. *Oncotarget* **7,** 71390–71399 (2016).

281. Quélard, D. *et al.* A cryptic frizzled module in cell surface collagen 18 inhibits Wnt/beta-catenin signaling. *PLoS One* **3,** e1878 (2008).

282. Moilanen, J. M. *et al.* Significant Role of Collagen XVII And Integrin β4 in Migration and Invasion of The Less Aggressive Squamous Cell Carcinoma Cells. *Sci. Rep.* **7,** 45057 (2017).

283. Kjellén, L. & Lindahl, U. Proteoglycans: Structures and Interactions. *Annu. Rev. Biochem.* **60,** 443–475 (1991).

284. Couchman, J. R. & Pataki, C. A. An introduction to proteoglycans and their localization. *J. Histochem. Cytochem.* **60,** 885–97 (2012).

285. Junqueira, L. C. & Montes, G. S. Biology of collagen-proteoglycan interaction. *Arch. Histol. Jpn.* **46,** 589–629 (1983).

286. Wight, T. N., Toole, B. P. & Hascall, V. C. in *The Extracellular Matrix: an Overview* 147–195 (Springer Berlin Heidelberg, 2011). doi:10.1007/978-3-642-16555-9_5

287. Prydz, K. & Dalen, K. T. Synthesis and sorting of proteoglycans. *J. Cell Sci.* **113 Pt 2,** 193–205 (2000).

288. Dick, G., Akslen-Hoel, L. K., Grøndahl, F., Kjos, I. & Prydz, K. Proteoglycan synthesis and Golgi organization in polarized epithelial cells. *J. Histochem. Cytochem.* **60,** 926–35 (2012).

289. Wight, T. N., Kinsella, M. G. & Qwarnström, E. E. The role of proteoglycans in cell adhesion, migration and proliferation. *Curr. Opin. Cell Biol.* **4,** 793–801 (1992).

290. Iida, J., Meijne, A. M. L., Knutson, J. R., Furcht, L. T. & McCarthy, J. B. Cell surface chondroitin sulfate proteoglycans in tumor cell adhesion, motility and invasion. *Semin. Cancer Biol.* **7,** 155–162 (1996).

291. Ruoslahti, E. Structure and Biology of Proteoglycans. *Annu. Rev. Cell Biol.* **4,** 229–255 (1988).

292. Yanagishita, M. Function of proteoglycans in the extracellular matrix. *Acta Pathol. Jpn.* **43,** 283–93 (1993).

293. O'Connell, M. P. *et al.* Heparan sulfate proteoglycan modulation of Wnt5A signal transduction in metastatic melanoma cells. *J. Biol. Chem.* **284,** 28704–12 (2009).

294. Nybakken, K. & Perrimon, N. Heparan sulfate proteoglycan modulation of developmental signaling in Drosophila.

295. Melrose, J., Hayes, A. J., Whitelock, J. M. & Little, C. B. Perlecan, the 'jack of all trades' proteoglycan of cartilaginous weight-bearing connective tissues. *BioEssays* **30,** 457–469 (2008).

296. Ricciardelli, C., Sakko, A. J., Ween, M. P., Russell, D. L. & Horsfall, D. J. The biological role and regulation of versican levels in cancer. *Cancer Metastasis Rev.* **28,** 233–245 (2009).

297. Sluiter, N. R. *et al.* Versican and vascular endothelial growth factor expression levels in peritoneal metastases from colorectal cancer are associated with survival after cytoreductive surgery and hyperthermic intraperitoneal chemotherapy. *Clin. Exp. Metastasis* **33,** 297–307 (2016).

298. Du, W. W., Yang, W. & Yee, A. J. Roles of versican in cancer biology--tumorigenesis, progression and metastasis. *Histol. Histopathol.* **28,** 701–13 (2013).

299. Seya, T. *et al.* Lumican expression in advanced colorectal cancer with nodal metastasis correlates with poor prognosis. *Oncol. Rep.* **16,** 1225–30 (2006).

300. Gu, X. *et al.* Up-regulated biglycan expression correlates with the malignancy in human colorectal cancers. *Clin. Exp. Med.* **12,** 195–199 (2012).

301. de Wit, M. *et al.* Lumican and versican protein expression are associated with colorectal adenoma-to-carcinoma progression. *PLoS One* **12,** e0174768 (2017).

302. Bi, X. *et al.* Oncogenic activin C interacts with decorin in colorectal cancer in vivo and in vitro. *Mol. Carcinog.* **55,** 1786–1795 (2016).

303. Andrlová, H. *et al.* Biglycan expression in the melanoma microenvironment promotes invasiveness via increased tissue stiffness inducing integrin-β1 expression. *Oncotarget* **8,** 42901–42916 (2017).

304. Ungefroren, H., Groth, S., Ruhnke, M., Kalthoff, H. & Fändrich, F. Transforming Growth Factor-β (TGF-β) Type I Receptor/ALK5-dependent Activation of the GADD45β Gene Mediates the Induction of Biglycan Expression by TGF-β. *J. Biol. Chem.* **280,** 2644–2652 (2005).

305. Lee, H.-C. *et al.* Cancer metastasis and EGFR signaling is suppressed by amiodarone-induced versican V2. *Oncotarget* **6,** 42976–87 (2015).

306. Li, F., Li, S. & Cheng, T. TGF-༂1 Promotes Osteosarcoma Cell Migration and Invasion Through the miR-143-Versican Pathway. *Cell. Physiol. Biochem.* **34,** 2169–2179 (2014).

307. Fanhchaksai, K. *et al.* Host stromal versican is essential for cancer-associated fibroblast function to inhibit cancer growth. *Int. J. Cancer* **138,** 630–641 (2016).

308. Wang, X. *et al.* Cancer-associated fibroblast-derived Lumican promotes gastric cancer progression via the integrin β1-FAK signaling pathway. *Int. J. Cancer* **141,** 998–1010 (2017).

309. Dawoody Nejad, L., Biglari, A., Annese, T. & Ribatti, D. Recombinant fibromodulin and decorin effects on NF-κB and TGFβ1 in the 4T1 breast cancer cell line. *Oncol. Lett.* **13,** 4475–4480 (2017).

310. Boivin, W. A. *et al.* Granzyme B Cleaves Decorin, Biglycan and Soluble Betaglycan, Releasing Active Transforming Growth Factor-β1. *PLoS One* **7,** e33163 (2012).

311. Jane D'Ardenne, A. & O'D. Mcgee, J. Fibronectin in disease. *J. Pathol.* **142,** 235–251 (1984).

312. Xiang, L. *et al.* The Extra Domain A of Fibronectin Increases VEGF-C Expression in Colorectal Carcinoma Involving the PI3K/AKT Signaling Pathway. *PLoS One* **7,** e35378 (2012).

313. Ou, J. *et al.* Endothelial cell-derived fibronectin extra domain A promotes colorectal cancer metastasis via inducing epithelial–mesenchymal transition. *Carcinogenesis* **35,** 1661–1670 (2014).

314. Bordeleau, F. *et al.* Tissue stiffness regulates serine/arginine-rich protein-mediated splicing of the extra domain B-fibronectin isoform in tumors. *Proc. Natl. Acad. Sci.* **112,** 8314–8319 (2015).

315. Hashimoto-Uoshima, M., Yan, Y. Z., Schneider, G. & Aukhil, I. The alternatively spliced domains EIIIB and EIIIA of human fibronectin affect cell adhesion and spreading. *J. Cell Sci.* **110 (Pt 18),** 2271–80 (1997).

316. Cai, X. *et al.* Down-regulation of FN1 inhibits colorectal carcinogenesis by suppressing proliferation, migration and invasion. *J. Cell. Biochem.* (2017). doi:10.1002/jcb.26651

317. Li, M. *et al.* Proteomic analysis of stromal proteins in different stages of colorectal cancer establishes Tenascin-C as a stromal biomarker for colorectal cancer metastasis. *Oncotarget* **7,** 37226–37237 (2016).

318. Murakami, T. *et al.* Tenascin C in colorectal cancer stroma is a predictive marker for liver metastasis and is a potent target of miR-198 as identified by microRNA analysis. *Br. J. Cancer* **117,** 1360–1370 (2017).

319. Takahashi, Y. *et al.* Tumor-derived tenascin-C promotes the epithelial-mesenchymal transition in colorectal cancer cells. *Anticancer Res.* **33,** 1927–34 (2013).

320. Gazzaniga, P. *et al.* Tenascin C and epidermal growth factor receptor as markers of circulating tumoral cells in bladder and colon cancer. *Oncol. Rep.* **14,** 1199–202 (2005).

321. DE WEVER, O. *et al.* Tenascin-C and SF/HGF produced by myofibroblasts in vitro provide convergent pro-invasive signals to human colon cancer cells through RhoA and Rac. *FASEB J.* **18,** 1016–1018 (2004).

322. Beiter, K. *et al.* β-Catenin regulates the expression of tenascin-C in human colorectal tumors. *Oncogene* **24,** 8200–8204 (2005).

323. Sarkar, S. *et al.* Activation of NOTCH Signaling by Tenascin-C Promotes Growth of Human Brain Tumor-Initiating Cells. *Cancer Res.* **77,** 3231–3243 (2017).

324. Rupp, T. *et al.* Tenascin-C Orchestrates Glioblastoma Angiogenesis by Modulation of Pro- and Anti-angiogenic Signaling. *Cell Rep.* **17,** 2607–2619

(2016).

325. Fazilaty, H., Gardaneh, M., Akbari, P., Zekri, A. & Behnam, B. SLUG and SOX9 Cooperatively Regulate Tumor Initiating Niche Factors in Breast Cancer. *Cancer Microenviron.* **9,** 71–74 (2016).

326. AUMAILLEY, M. *et al.* A simplified laminin nomenclature. *Matrix Biol.* **24,** 326–332 (2005).

327. Huang, D., Du, C., Ji, D., Xi, J. & Gu, J. Overexpression of LAMC2 predicts poor prognosis in colorectal cancer patients and promotes cancer cell proliferation, migration, and invasion. *Tumor Biol.* **39,** 101042831770584 (2017).

328. Lin, Q. *et al.* Analysis of colorectal cancer glyco-secretome identifies laminin β-1 (LAMB1) as a potential serological biomarker for colorectal cancer. *Proteomics* **15,** 3905–3920 (2015).

329. Fukazawa, S. *et al.* Laminin 3 expression as a prognostic factor and a predictive marker of chemoresistance in colorectal cancer. *Jpn. J. Clin. Oncol.* **45,** 533–40 (2015).

330. Guess, C. M., LaFleur, B. J., Weidow, B. L. & Quaranta, V. A Decreased Ratio of Laminin-332 3 to 2 Subunit mRNA is Associated with Poor Prognosis in Colon Cancer. *Cancer Epidemiol. Biomarkers Prev.* **18,** 1584–1590 (2009).

331. Lu, C.-L. *et al.* Inhibition of human 67-kDa laminin receptor sensitizes multidrug resistance colon cancer cell line SW480 for apoptosis induction. *Tumor Biol.* **37,** 1319–1325 (2016).

332. Kumazoe, M. *et al.* 67-kDa laminin receptor increases cGMP to induce cancer-selective apoptosis. *J. Clin. Invest.* **123,** 787–99 (2013).

333. Theocharis, A. D., Skandalis, S. S., Gialeli, C. & Karamanos, N. K. Extracellular matrix structure. *Adv. Drug Deliv. Rev.* **97,** 4–27 (2016).

334. Wessel, D. & Flügge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138,** 141–3 (1984).

335. Ishihama, Y. *et al.* Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **4,** 1265–72 (2005).

336. Qi, D. *et al.* A Software Toolkit and Interface for Performing Stable Isotope Labeling and Top3 Quantification Using Progenesis LC-MS. *Omi. A J. Integr. Biol.* **16,** 489–495 (2012).

337. Kim, S. *et al.* Carcinoma-produced factors activate myeloid cells through TLR2 to stimulate metastasis. *Nature* **457,** 102–106 (2009).

338. Krüger, A. Premetastatic niche formation in the liver: emerging mechanisms and mouse models. *J. Mol. Med.* **93,** 1193–1201 (2015).

339. Massagué, J. TGFβ signalling in context. *Nat. Rev. Mol. Cell Biol.* **13,** 616–630 (2012).

340. Gao, H. *et al.* Multi-organ Site Metastatic Reactivation Mediated by Non-canonical Discoidin Domain Receptor 1 Signaling. *Cell* **166,** 47–62 (2016).

341. Bresnick, A. R., Weber, D. J. & Zimmer, D. B. S100 proteins in cancer. *Nat. Rev. Cancer* **15,** 96–109 (2015).

342. M. Ferrão, P. *et al.* Inhibition of TGF-β pathway reverts extracellular matrix remodeling in T. cruzi -infected cardiac spheroids. *Exp. Cell Res.* **362,** 260–267 (2018).

343. Isella, C. *et al.* Stromal contribution to the colorectal cancer transcriptome. *Nat. Genet.* **47,** 312–319 (2015).

344. S.Y., L., A.E., Y. & A.N., G.-W. Tumor-infiltrating monocytes/macrophages promote tumor invasion and migration by upregulating S100A8 and S100A9 expression in cancer cells. in *Oncogene* **35,** 5735–5745 (Nature Publishing Group (Houndmills, Basingstoke, Hampshire RG21 6XS, United Kingdom), 2016).

345. Zhao, L. *et al.* Recruitment of a myeloid cell subset (CD11b/Gr1 ^{mid}) via CCL2/CCR2 promotes the development of colorectal cancer liver metastasis*. *Hepatology* **57,** 829–839 (2013).

346. Lim, S. Y., Yuzhalin, A. E., Gordon-Weeks, A. N. & Muschel, R. J. Tumor-infiltrating monocytes/macrophages promote tumor invasion and migration by upregulating S100A8 and S100A9 expression in cancer cells. in *Oncogene* **35,** 5735–5745 (2016).

347. Mederacke, I., Dapito, D. H., Affò, S., Uchinami, H. & Schwabe, R. F. High-yield and high-purity isolation of hepatic stellate cells from normal and fibrotic mouse livers. *Nat. Protoc.* **10,** 305–315 (2015).

348. Gressner, A. M. Transdifferentiation of hepatic stellate cells (Ito cells) to myofibroblasts: a key event in hepatic fibrogenesis. *Kidney Int. Suppl.* **54,** S39-45 (1996).

349. Lua, I. *et al.* Characterization of hepatic stellate cells, portal fibroblasts, and mesothelial cells in normal and fibrotic livers. *J. Hepatol.* **64,** 1137–1146 (2016).

350. de Jong, P. R., Mo, J.-H., Harris, A. R., Lee, J. & Raz, E. STAT3: An Anti-Invasive Factor in Colorectal Cancer? *Cancers (Basel).* **6,** 1394–407 (2014).

351. Banerjee, K. & Resat, H. Constitutive activation of STAT3 in breast cancer cells: A review. *Int. J. Cancer* **138,** 2570–2578 (2016).

352. Pópulo, H., Lopes, J. M. & Soares, P. The mTOR signalling pathway in human cancer. *Int. J. Mol. Sci.* **13,** 1886–918 (2012).

353. Azmi, A. S., Bao, B. & Sarkar, F. H. Exosomes in cancer development, metastasis, and drug resistance: a comprehensive review. *Cancer Metastasis Rev.* **32,** 623–42 (2013).

354. Calon, A. *et al.* Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat. Genet.* **47,** 320–329 (2015).

355. Rojas, A. *et al.* The aberrant methylation of *TSP1* suppresses TGF-β1 activation

in colorectal cancer. *Int. J. Cancer* **123,** 14–21 (2008).

356. Saharinen, J., Hyytiäinen, M., Taipale, J. & Keski-Oja, J. Latent transforming growth factor-beta binding proteins (LTBPs)--structural extracellular matrix proteins for targeting TGF-beta action. *Cytokine Growth Factor Rev.* **10,** 99–117 (1999).

357. Sipilä, K. H. *et al.* Extracellular citrullination inhibits the function of matrix associated TGF-β. *Matrix Biol.* **55,** 77–89 (2016).

358. David, C. J. *et al.* TGF-β Tumor Suppression through a Lethal EMT. *Cell* **164,** 1015–1030 (2016).

359. Roskams, T. *et al.* Heparan sulfate proteoglycan expression in normal human liver. *Hepatology* **21,** 950–8 (1995).

360. Cortés, V. & Rigotti, A. Liver heparan sulfate proteoglycans: Old molecules provide new insights on lipoprotein metabolism. *Hepatology* **45,** 1078–1080 (2007).

361. Kircher, M. & Kelso, J. High-throughput DNA sequencing - concepts and limitations. *BioEssays* **32,** 524–536 (2010).

362. Booth, C. M. & Eisenhauer, E. A. Progression-free survival: meaningful or simply measurable? *J. Clin. Oncol.* **30,** 1030–3 (2012).

363. Xing, F., Saidou, J. & Watabe, K. Cancer associated fibroblasts (CAFs) in tumor microenvironment. *Front. Biosci. (Landmark Ed.* **15,** 166–79 (2010).

364. Madar, S., Goldstein, I. & Rotter, V. 'Cancer associated fibroblasts'--more than meets the eye. *Trends Mol. Med.* **19,** 447–53 (2013).

365. P., C. Cancer-associated-fibroblasts and tumour cells: A diabolic liaison driving cancer progression. in *Cancer and Metastasis Reviews* **31,** 195–208 (Springer Netherlands (Van Godewijckstraat 30, Dordrecht 3311 GZ, Netherlands), 2012).

366. Kalluri, R. The biology and function of fibroblasts in cancer. *Nat. Rev. Cancer* **16,** 582–598 (2016).

367. Kalluri, R. & Zeisberg, M. Fibroblasts in cancer. *Nat. Rev. Cancer* **6,** 392–401 (2006).

368. Netchiporouk, E. *et al.* Deregulation in STAT signaling is important for cutaneous T-cell lymphoma (CTCL) pathogenesis and cancer progression. *Cell Cycle* **13,** 3331–5 (2014).

369. Guertin, D. A. & Sabatini, D. M. Defining the Role of mTOR in Cancer. *Cancer Cell* **12,** 9–22 (2007).

370. Francipane, M. G. & Lagasse, E. mTOR pathway in colorectal cancer: an update. *Oncotarget* **5,** 49–66 (2014).

371. Chakraborty, D. *et al.* Activation of STAT3 integrates common profibrotic pathways to promote fibroblast activation and tissue fibrosis. *Nat. Commun.* **8,** 1130 (2017).

372. Xiang, D.-M. *et al.* The HLF/IL-6/STAT3 feedforward circuit drives hepatic

stellate cell activation to promote liver fibrosis. *Gut* gutjnl-2016-313392 (2017). doi:10.1136/gutjnl-2016-313392

373. Pang, M. *et al.* A novel STAT3 inhibitor, S3I-201, attenuates renal interstitial fibroblast activation and interstitial fibrosis in obstructive nephropathy. *Kidney Int.* **78,** 257–268 (2010).

374. Prêle, C. M., Yao, E., O'Donoghue, R. J. J., Mutsaers, S. E. & Knight, D. A. STAT3. *Proc. Am. Thorac. Soc.* **9,** 177–182 (2012).

375. Woodcock, H. *et al.* mTOR signalling is an essential pathway for TGF-β $_1$ induced collagen synthesis. in *3.3 Mechanisms of Lung Injury and Repair* **46,** PA935 (European Respiratory Society, 2015).

376. Jiang, L. *et al.* Rheb/mTORC1 Signaling Promotes Kidney Fibroblast Activation and Fibrosis. *J. Am. Soc. Nephrol.* **24,** 1114–1126 (2013).

377. Yu, S.-Y., Liu, L., Li, P. & Li, J. Rapamycin Inhibits the mTOR/p70S6K Pathway and Attenuates Cardiac Fibrosis in Adriamycin-induced Dilated Cardiomyopathy. *Thorac. Cardiovasc. Surg.* **61,** 223–228 (2012).

378. Simler, N. R. *et al.* The rapamycin analogue SDZ RAD attenuates bleomycin-induced pulmonary fibrosis in rats. *Eur. Respir. J.* **19,** 1124–7 (2002).

379. Navé, B. T., Ouwens, M., Withers, D. J., Alessi, D. R. & Shepherd, P. R. Mammalian target of rapamycin is a direct target for protein kinase B: identification of a convergence point for opposing effects of insulin and amino-acid deficiency on protein translation. *Biochem. J.* **344 Pt 2,** 427–31 (1999).

380. Guasch-Ferré, M. *et al.* Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis. *Diabetes Care* **39,** 833–846 (2016).

381. de Bruin, C. & Dauber, A. Insights from exome sequencing for endocrine disorders. *Nat. Rev. Endocrinol.* **11,** 455–464 (2015).

382. Johar, A. S. *et al.* Candidate gene discovery in autoimmunity by using extreme phenotypes, next generation sequencing and whole exome capture. *Autoimmun. Rev.* **14,** 204–209 (2015).

383. Norman, P. J. *et al.* Defining KIR and HLA Class I Genotypes at Highest Resolution via High-Throughput Sequencing. *Am. J. Hum. Genet.* **99,** 375–391 (2016).

384. Qiao, D. *et al.* Exome Sequencing Analysis in Severe, Early-Onset Chronic Obstructive Pulmonary Disease. *Am. J. Respir. Crit. Care Med.* **193,** 1353–1363 (2016).

385. Laurila, P.-P. *et al.* Genomic, Transcriptomic, and Lipidomic Profiling Highlights the Role of Inflammation in Individuals With Low High-density Lipoprotein Cholesterol. *Arterioscler. Thromb. Vasc. Biol.* **33,** 847–857 (2013).

386. Würtz, P. *et al.* High-throughput quantification of circulating metabolites improves prediction of subclinical atherosclerosis. *Eur. Heart J.* **33,** 2307–16 (2012).

387. Khurana, R. *et al.* Identification of urinary exosomal noncoding RNAs as novel

biomarkers in chronic kidney disease. *RNA* **23,** 142–152 (2017).

388. Kalim, S. & Rhee, E. P. An overview of renal metabolomics. *Kidney Int.* **91,** 61–69 (2017).

389. Chen-Plotkin, A. S. Unbiased approaches to biomarker discovery in neurodegenerative diseases. *Neuron* **84,** 594–607 (2014).

390. Parikshak, N. N., Gandal, M. J. & Geschwind, D. H. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.* **16,** 441–458 (2015).

391. Dimitrakopoulos, L., Prassas, I., Diamandis, E. P. & Charames, G. S. Onco-proteogenomics: Multi-omics level data integration for accurate phenotype prediction. *Crit. Rev. Clin. Lab. Sci.* **54,** 414–432 (2017).

392. Pauli, C. *et al.* Personalized *In Vitro* and *In Vivo* Cancer Models to Guide Precision Medicine. *Cancer Discov.* **7,** 462–477 (2017).

393. Friedman, A. A., Letai, A., Fisher, D. E. & Flaherty, K. T. Precision medicine for cancer with next-generation functional diagnostics. *Nat. Rev. Cancer* **15,** 747–756 (2015).

394. Berg, E. L. Systems biology in drug discovery and development. *Drug Discov. Today* **19,** 113–125 (2014).

395. de Chassey, B., Meyniel-Schicklin, L., Vonderscher, J., André, P. & Lotteau, V. Virus-host interactomics: new insights and opportunities for antiviral drug discovery. *Genome Med.* **6,** 115 (2014).

396. O'Reilly, L. P. *et al.* A genome-wide RNAi screen identifies potential drug targets in a C. elegans model of α1-antitrypsin deficiency. *Hum. Mol. Genet.* **23,** 5123–5132 (2014).

397. Woo, J. H. *et al.* Elucidating Compound Mechanism of Action by Network Perturbation Analysis. *Cell* **162,** 441–451 (2015).

398. Backus, K. M. *et al.* Proteome-wide covalent ligand discovery in native biological systems. *Nature* **534,** 570–574 (2016).

399. Patel, P. S., Telang, S. D., Rawal, R. M. & Shah, M. H. A review of proteomics in cancer research. *Asian Pac. J. Cancer Prev.* **6,** 113–7

400. Thiery, J. P. & Sleeman, J. P. Complex networks orchestrate epithelial–mesenchymal transitions. *Nat. Rev. Mol. Cell Biol.* **7,** 131–142 (2006).

401. Yu, M. *et al.* Circulating Breast Tumor Cells Exhibit Dynamic Changes in Epithelial and Mesenchymal Composition. *Science (80-.).* **339,** 580–584 (2013).

402. Huang, R. Y.-J. *et al.* An EMT spectrum defines an anoikis-resistant and spheroidogenic intermediate mesenchymal state that is sensitive to e-cadherin restoration by a src-kinase inhibitor, saracatinib (AZD0530). *Cell Death Dis.* **4,** e915 (2013).

403. Nieto, M. A., Huang, R. Y.-J., Jackson, R. A. & Thiery, J. P. EMT: 2016. *Cell* **166,** 21–45 (2016).

404. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial–mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **15,** 178–196 (2014).

405. Tseng, J. H. *et al.* miR-200c-driven Mesenchymal-To-Epithelial Transition is a Therapeutic Target in Uterine Carcinosarcomas. *Sci. Rep.* **7,** 3614 (2017).

406. Slabáková, E., Culig, Z., Remšík, J. & Souček, K. Alternative mechanisms of miR-34a regulation in cancer. *Cell Death Dis.* **8,** e3100 (2017).

407. Thiery, J. P., Acloque, H., Huang, R. Y. J. & Nieto, M. A. Epithelial-Mesenchymal Transitions in Development and Disease. *Cell* **139,** 871–890 (2009).

408. Carver, E. A., Jiang, R., Lan, Y., Oram, K. F. & Gridley, T. The Mouse Snail Gene Encodes a Key Regulator of the Epithelial-Mesenchymal Transition. *Mol. Cell. Biol.* **21,** 8184–8188 (2001).

409. Veltmaat, J. M. *et al.* Snail is an immediate early target gene of parathyroid hormone related peptide signaling in parietal endoderm formation. *Int. J. Dev. Biol.* **44,** 297–307 (2000).

410. Kokkinos, M. I., Murthi, P., Wafai, R., Thompson, E. W. & Newgreen, D. F. Cadherins in the human placenta – epithelial–mesenchymal transition (EMT) and placental development. *Placenta* **31,** 747–755 (2010).

411. Saunders, L. R. & McClay, D. R. Sub-circuits of a gene regulatory network control a developmental epithelial-mesenchymal transition. *Development* **141,** 1503–1513 (2014).

412. Wu, S.-Y., Ferkowicz, M. & McClay, D. R. Ingression of primary mesenchyme cells of the sea urchin embryo: A precisely timed epithelial mesenchymal transition. *Birth Defects Res. Part C Embryo Today Rev.* **81,** 241–252 (2007).

413. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* **119,** 1420–8 (2009).

414. Zeisberg, E. M. *et al.* Endothelial-to-mesenchymal transition contributes to cardiac fibrosis. *Nat. Med.* **13,** 952–961 (2007).

415. Strutz, F. *et al.* Role of basic fibroblast growth factor-2 in epithelial-mesenchymal transformation. *Kidney Int.* **61,** 1714–1728 (2002).

416. Zeisberg, M. *et al.* Bone morphogenic protein-7 inhibits progression of chronic renal fibrosis associated with two genetic mouse models. *Am. J. Physiol. Physiol.* **285,** F1060–F1067 (2003).

417. Yang, J. *et al.* Disruption of tissue-type plasminogen activator gene in mice reduces renal interstitial fibrosis in obstructive nephropathy. *J. Clin. Invest.* **110,** 1525–1538 (2002).

418. Mani, S. A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133,** 704–15 (2008).

419. Morel, A.-P. *et al.* Generation of Breast Cancer Stem Cells through Epithelial-Mesenchymal Transition. *PLoS One* **3,** e2888 (2008).

420. Dalerba, P., Cho, R. W. & Clarke, M. F. Cancer Stem Cells: Models and

Concepts. *Annu. Rev. Med.* **58,** 267–284 (2007).

421. Tsai, J. H. & Yang, J. Epithelial-mesenchymal plasticity in carcinoma metastasis. *Genes Dev.* **27,** 2192–206 (2013).

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

423. Vega, S. *et al.* Snail blocks the cell cycle and confers resistance to cell death. *Genes Dev.* **18,** 1131–1143 (2004).

424. Eckert, M. A. *et al.* Twist1-Induced Invadopodia Formation Promotes Tumor Metastasis. *Cancer Cell* **19,** 372–386 (2011).

425. Stoletov, K. *et al.* Visualizing extravasation dynamics of metastatic tumor cells. *J. Cell Sci.* **123,** 2332–41 (2010).

426. Wendt, M. K., Balanis, N., Carlin, C. R. & Schiemann, W. P. STAT3 and epithelial-mesenchymal transitions in carcinomas. *JAK-STAT* **3,** e28975 (2014).

427. Wellner, U. *et al.* The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nat. Cell Biol.* **11,** 1487–1495 (2009).

428. Pegoraro, S. *et al.* HMGA1 promotes metastatic processes in basal-like breast cancer regulating EMT and stemness. *Oncotarget* **4,** 1293–308 (2013).

429. Brabletz, T. Cancer Cell Previews EMT and MET in Metastasis: Where Are the Cancer Stem Cells? *CCELL* **22,** 699–701 (2012).

430. Ocaña, O. H. *et al.* Metastatic Colonization Requires the Repression of the Epithelial-Mesenchymal Transition Inducer Prrx1. *Cancer Cell* **22,** 709–724 (2012).

431. Tsai, J. H., Donaher, J. L., Murphy, D. A., Chau, S. & Yang, J. Spatiotemporal Regulation of Epithelial-Mesenchymal Transition Is Essential for Squamous Cell Carcinoma Metastasis. *Cancer Cell* **22,** 725–736 (2012).

432. Lugli, E. B. *et al.* Expression of citrulline and homocitrulline residues in the lungs of non-smokers and smokers: implications for autoimmunity in rheumatoid arthritis. *Arthritis Res. Ther.* **17,** 9 (2015).

433. Sofat, N., Wait, R., Robertson, S. D., Baines, D. L. & Baker, E. H. Interaction between extracellular matrix molecules and microbial pathogens: evidence for the missing link in autoimmunity with rheumatoid arthritis as a disease model. *Front. Microbiol.* **5,** 783 (2015).

434. Clancy, K. W., Weerapana, E. & Thompson, P. R. Detection and identification of protein citrullination in complex biological systems. *Curr. Opin. Chem. Biol.* **30,** 1–6 (2016).

435. Liu, Y.-L., Chiang, Y.-H., Liu, G.-Y. & Hung, H.-C. Functional Role of Dimerization of Human Peptidylarginine Deiminase 4 (PAD4). *PLoS One* **6,** e21314 (2011).

436. Li, J. *et al.* Exosomes mediate the cell-to-cell transmission of IFN-α-induced antiviral activity. *Nat. Immunol.* **14,** 793–803 (2013).

437. Andrei, C. *et al.* The secretory route of the leaderless protein interleukin 1beta involves exocytosis of endolysosome-related vesicles. *Mol. Biol. Cell* **10,** 1463–75 (1999).

438. Sulzmaier, F. J., Jean, C. & Schlaepfer, D. D. FAK in cancer: mechanistic findings and clinical applications. *Nat. Rev. Cancer* **14,** 598–610 (2014).

439. Cox, T. R. & Erler, J. T. Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. *Dis. Model. Mech.* **4,** 165–78 (2011).

440. Lu, P., Weaver, V. M. & Werb, Z. The extracellular matrix: a dynamic niche in cancer progression. *J. Cell Biol.* **196,** 395–406 (2012).

441. Manfredi, S. *et al.* Epidemiology and Management of Liver Metastases From Colorectal Cancer. *Ann. Surg.* **244,** 254–259 (2006).

442. Hackl, C. *et al.* A population-based analysis on the rate and surgical management of colorectal liver metastases in Southern Germany. *Int. J. Colorectal Dis.* **26,** 1475–1481 (2011).

443. Stadler, S. C. *et al.* Dysregulation of PAD4-mediated citrullination of nuclear GSK3 activates TGF- signaling and induces epithelial-to-mesenchymal transition in breast cancer cells. *Proc. Natl. Acad. Sci.* **110,** 11851–11856 (2013).

444. DUAN, Q., PANG, C., CHANG, N., ZHANG, J. & LIU, W. Overexpression of PAD4 suppresses drug resistance of NSCLC cell lines to gefitinib through inhibiting Elk1-mediated epithelial-mesenchymal transition. *Oncol. Rep.* **36,** 551–558 (2016).

445. Bicker, K. L. & Thompson, P. R. The protein arginine deiminases: Structure, function, inhibition, and disease. *Biopolymers* **99,** 155–163 (2012).

446. Nakashima, K., Hagiwara, T. & Yamada, M. Nuclear Localization of Peptidylarginine Deiminase V and Histone Deimination in Granulocytes. *J. Biol. Chem.* **277,** 49562–49568 (2002).

447. Chang, X. & Han, J. Expression of peptidylarginine deiminase type 4 (PAD4) in various tumors. *Mol. Carcinog.* **45,** 183–196 (2006).

448. Cantarino, N. *et al.* Downregulation of the Deiminase PADI2 Is an Early Event in Colorectal Carcinogenesis and Indicates Poor Prognosis. *Mol. Cancer Res.* **14,** 841–848 (2016).

449. Qu, Y., Franchi, L., Nunez, G. & Dubyak, G. R. Nonclassical IL-1 Secretion Stimulated by P2X7 Receptors Is Dependent on Inflammasome Activation and Correlated with Exosome Release in Murine Macrophages. *J. Immunol.* **179,** 1913–1925 (2007).

450. Daniels, M. & Brough, D. Unconventional Pathways of Secretion Contribute to Inflammation. *Int. J. Mol. Sci.* **18,** 102 (2017).

451. Sheng, W. Versican Mediates Mesenchymal-Epithelial Transition. *Mol. Biol.*

Cell **17,** 2009–2020 (2006).

452. Rojas, A. *et al.* The aberrant methylation ofTSP1suppresses TGF-β1 activation in colorectal cancer. *Int. J. Cancer* **123,** 14–21 (2008).

453. Saharinen, J. Latent transforming growth factor-Î² binding proteins (LTBPs)â□”structural extracellular matrix proteins for targeting TGF-Î² action. *Cytokine Growth Factor Rev.* **10,** 99–117 (1999).

454. Meng, F. & Wu, G. The rejuvenated scenario of epithelial–mesenchymal transition (EMT) and cancer metastasis. *Cancer Metastasis Rev.* **31,** 455–467 (2012).

455. Chaffer, C. L., San Juan, B. P., Lim, E. & Weinberg, R. A. EMT, cell plasticity and metastasis. *Cancer Metastasis Rev.* **35,** 645–654 (2016).

456. Zheng, X. *et al.* Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527,** 525–530 (2015).

457. Tsai, J. H. & Yang, J. Epithelial-mesenchymal plasticity in carcinoma metastasis. *Genes Dev.* **27,** 2192–2206 (2013).

458. Tsai, J. H., Donaher, J. L., Murphy, D. A., Chau, S. & Yang, J. Spatiotemporal Regulation of Epithelial-Mesenchymal Transition Is Essential for Squamous Cell Carcinoma Metastasis. *Cancer Cell* **22,** 725–736 (2012).

459. Ocaña, O. H. *et al.* Metastatic Colonization Requires the Repression of the Epithelial-Mesenchymal Transition Inducer Prrx1. *Cancer Cell* **22,** 709–724 (2012).

460. Hur, K. *et al.* MicroRNA-200c modulates epithelial-to-mesenchymal transition (EMT) in human colorectal cancer metastasis. *Gut* **62,** 1315–1326 (2012).

461. Wu, Y.-H., Chang, T.-H., Huang, Y.-F., Huang, H.-D. & Chou, C.-Y. COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. *Oncogene* **33,** 3432–3440 (2014).

462. Han, L. *et al.* LTBP2 is a prognostic marker in head and neck squamous cell carcinoma. *Oncotarget* **7,** 45052–45059 (2016).

463. Deryugina, E. I. *et al.* LTBP3 promotes early metastatic events during cancer cell dissemination. *Oncogene* (2018). doi:10.1038/s41388-017-0075-1

464. Jung, S.-N. *et al.* LAMB3 mediates metastatic tumor behavior in papillary thyroid cancer by regulating c-MET/Akt signals. *Sci. Rep.* **8,** 2718 (2018).

465. Naba, A. *et al.* Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver. *BMC Cancer* **14,** 518 (2014).

466. Maier, T., Güell, M. & Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **583,** 3966–3973 (2009).

467. Krøigård, A. B., Larsen, M. J., Thomassen, M. & Kruse, T. A. Molecular Concordance Between Primary Breast Cancer and Matched Metastases. *Breast J.* **22,** 420–430 (2016).

468. Bertucci, F. *et al.* Comparative genomic analysis of primary tumors and metastases in breast cancer. *Oncotarget* **7,** 27208–19 (2016).

469. Bronte, V. & Pittet, M. J. The spleen in local and systemic regulation of immunity. *Immunity* **39,** 806–18 (2013).

470. Mittal, V. K., Bhullar, J. S. & Jayant, K. Animal models of human colorectal cancer: Current status, uses and limitations. *World J. Gastroenterol.* **21,** 11854–61 (2015).

471. Reyes-Castillo, Z., Muñoz-Valle, J. F. & Llamas-Covarrubias, M. A. Clinical and immunological aspects of anti-peptidylarginine deiminase type 4 (anti-PAD4) autoantibodies in rheumatoid arthritis. *Autoimmun. Rev.* (2017). doi:10.1016/j.autrev.2017.11.023

SUPPLEMENTARY TABLES