From the Department of Microbiology,
Tumor and Cell Biology
Karolinska Institutet, Stockholm, Sweden

MICROENVIRONMENTAL CONTROL OF MALIGNANT GROWTH

Benedek Bozoky

Karolinska Institutet

Stockholm 2023

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2023

ISBN 978-91-8017-146-5
Cover illustration: “The tumor microenvironment mapped to the Karolinska Institute’s research microenvironment.” By Felicia Bozoky

Microenvironmental Control of Malignant Growth

Thesis for Doctoral Degree (Ph.D.)

By

Benedek Bozoky

The thesis will be defended in public at Franklinsalen, Tomtebodavägen 6, Solna, November 10th, 2023, at 9:30 am.

Principal Supervisor:
Professor Ingemar Ernberg
Karolinska Institute
Department of Microbiology, Tumor and Cell Biology

Co-supervisor(s):
Professor Laszlo Szekely
Karolinska Institute
Department of Laboratory Medicine
Division of Pathology

Professor Rickard Sandberg
Karolinska Institute
Department of Cell and Molecular Biology

Dr. Daniel Salamon
Karolinska Institute
Department of Women's and Children's Health

Opponent:
Professor Kristian Pietras
Lund University
Department of Laboratory Medicine
Division of Division of Translational Cancer Research

Examination Board:
Professor Carolina Wählby
Uppsala University
Department of Information Technology

Docent Jonas Fuxe
Karolinska Institute
Department of Laboratory Medicine
Division of Pathology

Dr. Nicola Crosetto
Karolinska Institute
Department of Department of Microbiology, Tumor and Cell Biology
Human Technopole, Milan
Department of Genomics Research Centre

There is no escape: It seems that everything people do to earn a livelihood, to subsist, or to enjoy life turns out to be illegal, immoral, or fattening, or - most disturbing - possibly carcinogenic.

- *Robbins Pathology - the concluding remarks of the chapter on cancer epidemiology*

To my family, friends, and patients.

Popular science summary of the thesis

Today, we understand that cancers can only form and spread through a complex interaction with their surrounding normal tissue. This process involves multiple components like the different cells of the immune system, fibroblasts that build and maintain the connective tissue, the extracellular matrix, and blood and lymphatic vessels. Replicating such complexity in a laboratory setting poses considerable challenges. Consequently, there exists a need to investigate these interactions in human tissue samples, where the full spectrum of tissue complexity is preserved.

We developed a method to analyze an online image database, the Human Protein Atlas (HPA), that contains information on what proteins are expressed in normal tissues and how they differ in the tissues surrounding tumors. With this method, we identified several proteins not previously known to be expressed in fibroblast surrounding tumors. Many of these were connected to the protein RhoA. To study its importance, we engineered fibroblasts that lacked this protein and measured how this affected cancer cell growth. We found that when fibroblasts can't express this protein, the cancer cells grow faster and form tumors more easily in mice, meaning that this protein most likely functions to help suppress cancers.

The molecule Decorin is typically found deposited in the stroma of normal tissues and has been described as having an inhibitory role. With the help of our method, we demonstrated its absence in the tissues surrounding tumors, thereby reinforcing its potential protective role in cancer.

Previously, research conducted on cell cultures has demonstrated that cancer cells can be influenced by their microenvironment and, to some extent, revert to a non-cancerous phenotype. However, the extent of this phenomenon in human tumors has not been thoroughly studied. To investigate this, we analyzed several cases of pancreatic cancer in which cancer cells infiltrate the small intestine. In this context, our findings revealed that depending on their specific location, cancer cells underwent a transition from a more aggressive form of pancreatic cancer to a more indolent one. They adopted characteristics resembling those of normal intestinal cells and exhibited altered rates of cell division. This highlights their plasticity to change due to influences from the tissue environment in which they reside.

Our method allowed us to analyze the HPA database; however, it still demanded substantial researcher input, rendering it time-consuming. To facilitate a more efficient exploration of the entire database, we developed an additional method based on artificial intelligence. Using this approach, we successfully demonstrated that a limited set of images of prostate basal cells enabled us to comprehensively search the database and identify similar ones. Employing this method, we identified 44 new markers associated with this cell type.

In summary, our methods have enabled the exploration of the alterations occurring around cancers and have improved our understanding of changes in the tumor microenvironment and its profound impact on cancer cell behavior and progression.

Abstract

The tumor microenvironment (TME) comprises a complex milieu of different cell types, including cancer associated fibroblasts (CAFs) and immune cells, blood vessels, and the extracellular matrix. Through its interaction with cancer cells, it plays an essential role in cancer invasion and metastasis. The inherent complexity of the TME presents a challenge to study it within experimental model systems. It underscores the importance of complementing such research with observation from human tumor tissues, wherein this intricate complexity is preserved.

In **Paper IV**, we introduce a new software designed to explore the Human Protein Atlas, an online database that includes image data on the protein expression across normal and cancerous tissues from immunohistochemically (IHC) stained tissues.

In **Paper I,** we use this software to identify 12 novel proteins expressed in cancer-associated fibroblasts, four revealing connections to Rho-kinase signaling. We contrast their expression across various tumors and against normal tissue fibroblasts, uncovering expression variability among cancer types and confirm their similarities with the myofibroblastic phenotype.

In **Paper II**, we explore the expression of the proteoglycan Decorin, abundantly present in normal connective tissue and having tumor inhibitory properties, showing its downregulation in the connective tissue surrounding tumors.

In **Paper III**, based on our observations in **Paper I** of the connection of Rho-signaling in CAFs, we study the effects of knocking out the related RhoA in fibroblasts both *in vitro* and *in vivo* models. We demonstrate that the knockout fibroblasts compromise their tumor inhibitory capacity, enhancing cancer cell growth, migration, and metastasis.

In **Paper VI**, we develop a new method for analyzing the extensive data within the Human Protein Atlas by developing a deep-learning-based image classifier. Utilizing a limited training image set, we classify all images available for the prostate, identifying 44 new markers of prostate basal cells.

In **Paper IV**, we explore the influence of the TME on cancer cells by systematically analyzing 20 pancreatic cancer patient samples utilizing an IHC panel. We define shifts in cancer cell phenotype relative to tissue localization, including a transition

to a more indolent cancer phenotype, an effect on cancer cell proliferation, and a tendency to normalize the cancer cell phenotype.

In conclusion, we developed two new methods that enable us to study protein expression in normal and cancerous tissues by enhancing the capabilities of the HPA. We identified new markers of CAFs and revealed a connection to Rho-signaling. Knocking out the related RhoA in experimental systems resulted in the fibroblasts losing their cancer inhibitory capacity. Finally, we show the remarkable plasticity of cancer cells, demonstrating that their phenotype undergoes significant alterations based on their spatial localization within normal tissue.

List of scientific papers

I. Novel Signatures of Cancer-associated Fibroblasts.
International Journal of Cancer 133, no. 2 (2013): 286-293.
Bozóky, B., Savchenko, A., Csermely, P., Korcsmáros, T., Dúl, Z., Pontén, F., Székely, L., & Klein, G.

II. Decreased Decorin Expression in the Tumor Microenvironment.
Cancer Medicine 3, no. 3 (2014): 485-491.
Bozoky, B., Savchenko, A., Guven, H., Ponten, F., Klein, G,. & Szekely, L.

III. RhoA Knockout Fibroblasts Lose Tumor-inhibitory Capacity in Vitro and Promote Tumor Growth in Vivo.
Proceedings of the National Academy of Sciences 114, no. 8 (2017): E1413-E1421.
Alkasalias, T., Alexeyenko, A., Hennig, K., Danielsson, F., Lebbink, R. J., Fielden, M., Turunen, S. P., Lehti, K., Kashuba, V., Madapura, H. S., **Bozoky, B.,** Lundberg, E., Balland, M., Guvén, H., Klein, G., Gad, A. K., & Pavlova, T

IV. Stabilization of the classical phenotype upon integration of pancreatic cancer cells into the duodenal epithelium.
Neoplasia, *23*(12), (2021): 1300-1306.
Bozoky, B., Fernández Moro, C., Strell, C., Geyer, N., Heuchel, R. L., Löhr, J. M., Ernberg, I., Szekely, L., Gerling, M., & Bozóky, B.

V. AtlasGrabber: a software facilitating the high throughput analysis of the human protein atlas online database
BMC Bioinformatics, 23(1), (2022): 546.
Bozoky, B., Szekely, L., Ernberg, I., & Savchenko, A.

VI. Identification of novel protein markers of prostate basal cells by application of deep learning to images from the Human Protein Atlas
Bozoky, B., Szekely, L., Alexeyenko, A., Ernberg, I., Petrov, I.
Manuscript.

Contents

List of abbreviations

TME	Tumor microenvironment
CAF	Cancer associated fibroblast
IHC	Immunohistochemistry
ECM	Extracellular matrix
BM	Basement membrane
TGF-beta	Transforming growth factor-beta
MCP1	Monocyte chemoattractant protein 1
PDGF	Platelet derived growth factor
FGF	Fibroblast growth factor
SLRP	Small leucine-rich proteoglycan
RTK	Receptor tyrosine kinases
HGFR	Hepatocyte growth factor receptor
IGF1R	Insulin like growth factor receptor I
EMT	Epithelial to mesenchymal transition
MET	Mesenchymal to epithelial transition
IF	Immunofluorescence
HER2	Human epidermal growth factor receptor 2
ER	Estrogen receptor
PR	Progesterone receptor
FFPE	Formalin fixed paraffin embedded
HIER	Heat induced epitope retrieval
PIER	Proteolytic induced epitope retrieval
AP	Alkaline phosphatase
HRP	Horseradish peroxidase
DAB	3,3′-Diaminobenzidine
AEC	3-Amino-9-Ethylcarbazole
BCIP	5-Bromo-4-Chloro-3-Indolyl-Phosphate

NBT	Nitro Blue Tetrazolium
HPA	Human Protein Atlas
RNAseq	RNA sequencing
PrEST	Protein Epitope Signature Tag
RGB	Red, Green, and Blue
HSB	Hue, Saturation, Color
KO	Knockout
WT	Wild type
TIRF	Total Infernal Reflection Fluorescence
FACS	Fluorescence activated cell sorting
scRNAseq	Single-cell RNA sequencing
NEA	Network enrichment analysis
GO	Gene Ontology
DO	Disease Ontology
PDAC	Pancreatic ductal adenocarcinoma
HMGA2	High-motility group AT-hook 2
ACTA2	Smooth muscle alpha actin

Introduction

Cancer is the second leading cause of death globally, and it is estimated that 1 in 2 men and 1 in 3 women will develop cancer during their lifetime in the United States [1]. Meanwhile, as populations in emerging countries adopt cancer associated lifestyles and changing demographics with increased age, the number of cases is expected to increase by 70% over the next two decades. The economic costs are already high, estimated to cost over 50 billion Euros per year in Europe, taking the cost of care, treatment, and loss of productivity into account [2, 3].

One of the reasons for the difficulty of treating cancer is the enormous diversity that cancer covers. There exist hundreds of types and subtypes. They are diverse in the various cells and tissues of origin, their different stages, molecular and genetic subclasses, and patient-to-patient variations. It is now also recognized that surrounding and distant normal cells are recruited to the tumor, making up the tumor microenvironment, further adding to the complexity [4].

Several hallmarks were proposed as an organizing principle to characterize cancer cells to make sense of decades of research and to find commonality in the vast diversity of cancer[2, 5]. As new research emerged, it was later revised[6, 7] and now encompasses ten hallmarks: (1) resistance to cell death, (2) sustained proliferative signaling, (3) evasion of growth suppression, (4) activation of invasion and metastasis, (5) genome instability and mutation,(6) induction of angiogenesis, (7) deregulation of cellular metabolism, (8) tumor-promoting inflammation, (9) replicative immortality, and (10) avoidance of immune destruction.

As these traits are common to all cancers, they make a helpful basis for new cancer targets. Their understanding has led to some remarkable success as immune therapies against melanomas [8, 9] but most have failed to live up to expectations. Most are transiently efficient and work only in a subset of patients [10–12].

For increased effectiveness, it has been proposed that a logical combination of multiple targeted therapies against the hallmarks could work together and disrupt several characteristics, acting synergistically and complementary. The more hallmarks that can be disabled, the higher the chance of killing all the cancer cells, not allowing it to recur or develop resistance [2, 13]. One could, for instance, target angiogenesis and block invasion and metastasis at the same time. Such targeting of multiple hallmarks could be further enhanced by targeting individual hallmarks

with multiple drugs. For example, inhibition of proliferative signaling by targeting both the BRAF and its downstream MEK is used in melanoma to inhibit the EGFR pathway [14–16].

Many of the hallmarks are well understood and targeted therapies already exist for several of these characteristics [4, 6]. For others, more understanding and new targets are needed. One exciting field of research constitutes the normal cells recruited to the tumor microenvironment, which, like cancers, are highly heterogeneous and diverse, but their understanding could offer new potential therapeutic targets [2, 4].

1 Literature review

1.1 The tumor microenvironment

The importance of the tumor microenvironment (TME) (Figure 1) in cancer formation and progression is now well accepted. In the past, there was a substantial focus on studying cancer cells alone and how the loss or gain of oncogenes and tumor suppressor genes leads to transformation. It is now understood that cancer cells do not act alone, but they recruit and corrupt resident and distant normal cell types in their surroundings. It is accepted as a fundamental part of cancer biology, and understanding it is crucial for exploiting it for diagnostic and therapeutic purposes [4].

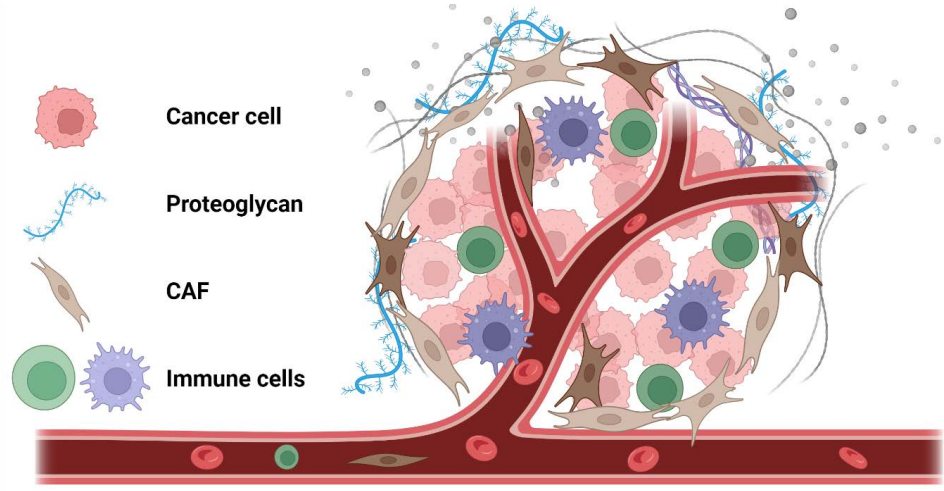

Figure 1. A simplified schematic representation of the tumor microenvironment. Its complexity is represented by the multiple components: cancer associated fibroblasts, immune cells, blood vessels, proteoglycans and blood vessels, and extracellular vesicles. Created with BioRender.com

One of the earlier evidences of the importance of the TME comes from observations where cancers develop due to chronic inflammation in which the normal tissue architecture is disrupted. Examples of this include cirrhosis of the liver from alcoholic liver disease and hepatitis that leads to an increased risk of hepatocellular cancer, gastric H. Pylori infection that can lead to gastric cancer, and inflammatory bowel disorders that result in an increased risk for colon cancer. Common in these is the chronic inflammation that results in a deranged environment and abnormal tissue environment [17].

It is also becoming apparent that most cancer hallmarks are maintained and aided by the contribution of stromal cells [4]. As we learn to understand it better, it can

provide treatment options, particularly in the context of the hallmarks as described above.

How the different components contribute to the hallmarks, what their functions are, and understating their diversity remain largely open questions yet to be answered [4]. Already, they have been implicated with proposed roles in resistance to therapy (radio, chemo, and targeted) where stromal cells protect cancer cells or subpopulations of them [4, 18].

1.2 Cancer-associated fibroblasts

Fibroblasts are one of the principal cells of the normal connective tissues with diverse, multifunctional roles. They deposit components of the extracellular matrix (ECM) and the basement membrane (BM), and take part in regulating cell differentiation, immune system modulation, and maintaining the homeostasis of tissues [19, 20].

In the tumor microenvironment (TME), they are typically present in larger numbers, differ from normal fibroblasts, and are known as cancer-associated fibroblasts (CAF). One study showed that initiated prostate epithelial cells, when co-injected with CAFs into mice, gave rise to tumors. At the same time, this did not occur when they were co-injected with normal fibroblasts [21]. Another study found that CAFs increased the number of metastases while normal fibroblasts had the opposite effect [22].

It is still unclear where these fibroblasts arise. The evidence points to the fact that they can originate from resident fibroblasts recruited from the BM [23] and studies indicate that they can originate from endothelial to mesenchymal transition [24] or from epithelial to mesenchymal transition in prostate and breast cancers [25, 26].

Several factors have also been proposed to play a role in the activation of CAFs. These include various growth factors and cytokines from the surrounding cells, in particular: TGF-beta (transforming growth factor beta), MCP1 (monocyte chemoattractant protein 1), PDGF (platelet derived growth factor), FGF (fibroblast growth factor), and different proteases [20, 27].

In turn, the activated CAFs will become a source of growth factors supporting the tumor's growth [28]. Other mechanisms have been proposed to help tumors, such as producing pro-inflammatory factors activating NF-kB signaling [29].

1.3 The extracellular matrix

The extracellular matrix (ECM) is a network of various molecules constituting the acellular environment. It has both biophysical and biochemical properties [30, 31] that are tissue-specific and made up of various proteins, glycoproteins, and proteoglycans. These often undergo complex transcriptional, translational, and post-translational modifications [32]. The components are often modular in that different rearrangements, and alternative splicing can combine a limited number of modules into many more functionally different molecules [33].

Abnormal ECM has been associated with diseases, including fibrosis [34], chronic inflammation [35] and cancer [36–38]. Genetic mutations of ECM components can result in severe illness, as in chondrodysplasias, osteogenesis imperfecta, and epidermolysis bullosa [39, 40].

In the tumor microenvironment, its composition has already been shown to contribute to clinical prognosis. One study looked at breast cancers and used ECM composition alone to stratify it into subclasses correlated with clinical prognosis. High expression of integrins and MMPs have been associated with poor prognosis, while high expression of protease inhibitors indicated good prognosis [41].

Critical steps to understanding the ECM's role in diseases will be the quantification and identification of its components in different diseases and normal states. Due to its essential functions and association with cancer prognosis, it is an attractive part of the TME to be studied.

1.4 Decorin

Decorin is an example of a molecule of the ECM that is believed to have an important role in cancer. It belongs to the small leucine-rich proteoglycan (SLRP) gene family and is predominantly made and deposited to the ECM by fibroblasts. It received its name from its attachment to collagen I fibers, which it “decorates” [42–44].

The core domain of decorin, termed decoron, spans 396 amino acids and features multiple leucine-rich repeats. A single decoron can bind up to six collagen fibers and is believed to have a key role in fibrillogenesis [45–48].

It has been described to have a multitude of functions in the ECM. It is believed to have a role in inflammation by binding toll-like receptors [49]. In addition, it can

directly antagonize receptor tyrosine kinases (RTK), for example, the hepatocyte growth factor receptor (HGFR), the insulin-like growth factor receptor I (IGF1R), and the epidermal growth factor receptor [50–53].

Its role in cancer has been termed the "guardian from the matrix" in a review due to its believed tumor suppressive role [54]. Studies have, for instance, shown that disrupting decorin expression can accelerate the formation of lymphomas and increase the formation of intestinal tumors in mouse models [55, 56]. While in vitro co-culture studies have shown that fibroblasts decrease their decorin expression when grown with cancer cells [57].

Studies on Decorin expression in the tumor microenvironment are contradictory, showing that its expression is decreased in lung cancer [58] and at various stages in breast cancer, with decreasing expression in more advanced cancer [59]. Other studies have, however, shown that its expression is increased in other cancers, such as Kaposi sarcomas or pancreatic cancers [60–62].

Various methods have been proposed to explain its observed tumor-suppressive nature, like the binding and downregulating RTKs like EGFR [63], suppressing β catenin levels [53] and thereby inhibiting cell growth and migration, and inhibiting angiogenesis [64]. An alternative method could be explained by the observation that naked collagen aids tumor growth, as seen in *in vitro* studies, and its absence from collagen could enable this in the tumor microenvironment [65–68].

1.5 Phenotype switching and cancer cell plasticity

For cancer cells to gain a migratory and invasive trait, they are suggested to undergo switches in their phenotype. An epithelial to mesenchymal transition (EMT) occurs when they lose their epithelial markers and gain mesenchymal traits, allowing them to migrate to distant tissues [69]. The secondary metastatic lesions will typically exhibit an epithelial-like phenotype, indicating transition back, a mesenchymal to epithelial transition (MET) of equal importance [70–72]. Both these transitions are proposed to arise partly from environmental cues; for example, increased TGF-beta signaling has been shown to induce EMT [73].

It is also known that cancer cells in the right context can be induced to differentiate and revert to resemble a normal phenotype. The earliest indication for this came from the work by Leo Sachs, who showed that malignancy could be reversed by inducing leukemic cells to differentiate [74]. Today, differentiation therapy is successfully used to treat acute promyelocytic leukemia [75, 76].

The influence of the tissue context was highlighted by co-culture experiments revealing that normal stromal cells can inhibit the progression of malignancy [77]. It was shown that when epithelial cells were grown in 3D gels containing extracellular matrix proteins, they regained their polarity, formed acinar structures, and suppressed oncogenes' transforming and proliferative features [78, 79].

Similarly, when human breast cancer cells were co-cultured in 3D laminin-rich gels, they reverted to a near normal phenotype [80] and additional in vitro experiments showed that certain tumors could differentiate and their growth can be suppressed by placing them in direct contact with an intact basement membrane [81–83]. These and similar observations indicate that the environment in which cells reside can significantly influence that cancer cell's phenotype, overriding their cancer genotype [84, 85].

Even early animal studies could show that the highly oncogenic Rous sarcoma virus is not carcinogenic in the early chicken embryo [86] But if the cells were taken from their normal environment and cultured on a dish, they transformed [87].

It is evident that cancer cells do not exist in isolation but interact closely with their surrounding tumor microenvironment. They take part in modifying it, often to their advantage, but in turn, they also receive cues from it that can profoundly influence their behavior.

Many interactions described above are thought to occur at a tumor's invasive edge, elsewhere described as the tumor margin or leading edge. It is the meeting place where cancer cells come in physical contact with the stromal, the ECM, and first contact with immune cells. This is where the CAFs provide the pro-invasive factors such as TGF-beta and PDGF signaling to promote EMT [88], and immune cells like macrophages are recruited by tumor-derived chemoattractants [89] and, in turn, supply migratory factors like EGF, take part in proteolytic remodeling of the ECM and regulate fibrillary collagen production, all to help promote tumor invasion [90–92].

The tumor microenvironment is highly diverse and complex. Many different cells and extracellular components exist in an interplay between them and the cancer cells. Studying this complex network poses challenges but could offer many potential benefits by revealing new targets for cancer treatment.

Most of today's cancer therapeutics focus on cancer cells. The stromal elements offer attractive alternatives as it is genetically stable and, therefore, less susceptible to developing therapeutic resistance. Only a few current oncological treatments have targeted the tumor microenvironment, but with some success, as in the angiogenesis inhibition [93]. Nevertheless, therapies aimed not at the destruction or depletion of specific components, but their modification has fared better. Immunotherapies are an example of re-education of the TME that is gaining ground with better and better results [13, 84].

Much of the evidence of the functions of stromal cells in a tumor context originates from experiments from model systems. These are typically either (1) genetically modified mouse models of cancers, (2) human xenotransplant mice, or (3) cell and organ co-culture assays [4]. These are limited in that they can only model the complexity of human tumors to a certain degree. They are, for instance, limited in studying the diverse spectrum of malignancies that exist in different stages, originate from different cells, are heterogeneous in their genetic and epigenetic derangements, and differ from patient to patient.

The path toward clarification is a challenge, but one proposed method, described in a review by D. Hanahan [4], is to attempt to integrate the model systems with biopsies or surgically resected specimens from human cancer patients. He proposes that these could be done by (1) advanced histochemical methods such as multicolor immunostaining and in situ hybridizations, (2) precise laser capture microdissection of cell types and subtypes from the tumor stroma, and (3) selective antibody capture by flow cytometry.

By combining different specific therapies, one can hope to develop better responses to cancer treatments. The hallmarks and underlying components should be well understood to exploit this fully. Much research is focused on studying the tumor microenvironment, and there are already attempts to target it. One of the limitations, however, comes from the model systems that are used to examine it. Current research can be aided by studying human tumors, with, for example, in situ hybridizations or immunohistochemistry. These studies have the potential to bridge knowledge from the model systems to human cancers and offer insights into questions that the model systems cannot hope to resolve.

1.6 Immunohistochemistry

1.6.1 Background and methodology

Immunohistochemistry (IHC) is a technique that originates from the combination of immunology and pathology. It is based on the antibodies binding to specific antigens in tissues and visualized by either an enzymatic reaction (chromogenic IHC) or immunofluorescence (IF). The technique was pioneered by the work of Albert Coons in 1941, who first showed how labeled antibodies could be used to identify specific antigens in tissues with immunofluorescence [94]. In this, the antibodies were labeled with a fluorescent dye that produces a light when excited by a laser and could be detected by a fluorescence microscope [95]. Later, in 1967, the work by Nakane and Pierce introduced the chromogenic method to visualize antibodies[96]. This method was based on an enzymatic reaction, which catalyzes a color-forming reaction that can be visualized with a light microscope.

The chromogenic IHC is widely used for both predictive and prognostic information in pathology, for example, testing for human epidermal growth factor receptor 2 (HER2) amplification [97], estrogen receptor (ER) and progesterone (PR) receptor detection in breast cancer [98, 99]. Another example is their use in subtyping cancers based on their tissue of origin. This is used in diagnosing cancers of unknown primary sites, where broad IHC panels are used to identify the tissue of origin so that appropriate treatments can be administered [100, 101].

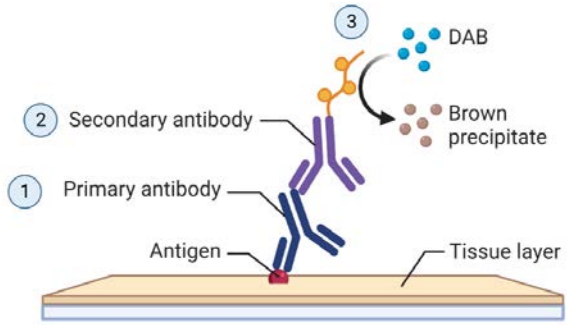

Figure 2. The process of IHC. 1) A primary antibody binds to the antigen of interest. 2) Typically, a secondary antibody is applied that binds the first antibody. 3) In the chromogenic method, an enzymatic reaction produces a color, here depicted using DAB, specific to the antigen sites, which can be visualized with light microscopy. Created with BioRender.com

IHC is typically applied to surgical specimens or biopsies that are formalin-fixed paraffin-embedded (FFPE), although it can also be used on frozen sections. The steps of IHC involve tissue preparation, antigen retrieval, the addition of a primary

antibody, the addition of a secondary antibody and detection, and counterstaining (Figure 2).

Proper tissue preparation is critical to ensure that the sample retains its tissue architecture, cell morphology, and the antigenicity of the target epitopes. It involves fixation, which chemically cross-links proteins to stop cellular processes, freezes the cellular components in place and in their current conformation, provides some structural support for the tissue, and prevents degradation. Commonly, FFPE formalin is used for this step [102–104]. However, this step can also be done using an alcohol solution or other aldehydes.

In the also commonly used frozen tissue sections, the tissue is first embedded in a cryoprotective medium and then frozen. Fixation is carried out after the sectioning process. Cryostats are used to section frozen tissues, which have the advantage of shorter processing times and better preservation of sensitive epitopes. However, in terms of preserving histological morphology, FFPE tissues are often superior to frozen tissues. In the case of FFPE, after the fixation, the tissues are embedded in paraffin blocks, which is necessary for sectioning but also supports preserving the tissue architecture and integrity and further helps preserve the tissue.

Antigen retrieval is necessary as the fixation method can mask epitopes and obstruct the detection of the target by the primary antibody. This process reverts the crosslinking of proteins. Commonly, this is done either by heat-induced epitope retrieval (HIER) or proteolytic-induced epitope retrieval (PIER), which is based on enzymatic digestion [105].

The primary antibodies are selected to target the epitope in question. Therefore, it is important that they are highly specific and can be reliably reproduced to ensure that they do not produce off-target binding. These antibodies can be either monoclonal, which binds a single, specific epitope, or polyclonal, binding several epitopes on the target. They differ in their production and have inherent advantages and disadvantages.

Monoclonal antibodies are specific for a single epitope and can be produced from immortalized B-cell lines [106, 107]. Because of this, they can be produced with higher consistency and a known specificity. This makes them more suitable for routine clinical pathology and for semi-quantitative analysis in the clinical setting. However, since they only recognize a single epitope, changes in the epitope from

protein isoforms or denaturation can limit their epitope binding. Selecting monoclonal antibodies is advantageous when high specificity, reproducibility, and consistency is necessary. Nevertheless, while monoclonal antibodies have a known epitope binding, their results can be challenging to interpret if their specificity is low or if the target epitope is not abundant enough or might be masked.

Polyclonal antibodies are produced by immunizing animals with the target antigen. They will result in a mixture of antibodies that bind against multiple epitopes on the target, as in a normal adaptive immune response. The advantages over monoclonal antibodies are that they have a higher tolerance for variability in the target and are better at detecting antigens that could undergo conformational changes. In addition, they can have a higher affinity to the antigen and, therefore, give a stronger signal. However, due to its polyclonality, there is an increased risk of off-target binding due to an increased likelihood of cross-reactivity with other epitopes. Another disadvantage is that due to their production method, their different batches will vary in their polyclonal setup; thus, reproducibility is challenging [108].

The detection can be done by either IF or enzymatic activity. An indirect detection method is typically used when an unlabeled (primary) antibody that binds the target epitope is targeted with a secondary antibody with a labeled enzyme [109]. This results in signal amplification due to the binding of several secondary antibodies to the primary one. Typically used enzyme substrates are alkaline phosphatase (AP) and horseradish peroxidase (HRP). Several different chromogenic substrates are also available, commonly (3,3'-Diaminobenzidine) DAB, 3-Amino-9-Ethylcarbazole (AEC), or 5-Bromo-4-Chloro-3-Indolyl-Phosphate (BCIP) / Nitro Blue Tetrazolium (NBT). A final counterstaining step is then applied, usually with a Hematoxylin staining. This final step helps to visualize the tissue architecture and cells in the non-target areas [110].

1.6.2 The interpretation of IHC stainings

Immunohistochemistry will provide information on whether an antibody is present in tissues and is, therefore, mainly considered a qualitative technique. One of its limitations is that it cannot be considered a purely quantitative method, meaning that it cannot be used to measure the amount of target protein present.

One reason for this is that antigen-antibody reactions are not considered stoichiometric, meaning that there is no defined proportional correlation

between the antibody binding to its target and, therefore, there is no clear correlation between staining strength and quantity of antigen present. The reason for this includes that the target antibodies can be accessible to different degrees in the tissue due to factors like antigen retrieval efficacy variations in epitope availability from, for example, post-translational modification and background binding of antibodies where they bind non-targets. In addition, IHC also involves a series of amplification steps, like using a secondary antibody to visualize the results, making it difficult to control the final intensity of the amplified signal in terms of the amount of antigen. In addition, the commonly used DAB staining does not adhere to the Beer-Lambert law, which describes the linear correlation between the concentration of a compound and its light absorbance. The reason is that the brown staining it produces does not absorb light but rather scatters it and has a broad light spectrum [111].

The qualitative interpretation of IHC can, however, to some extent, involve a quantitative step, where the number of positive cells is counted, as is used in the number of cancer cells expressing the ER or PR receptors in breast cancer [99], or nowadays widely used for the quantification of PD1 and PDL1 expression to identify patients suitable for immunotherapy in certain cancers [112–114].

Despite these limitations, IHC is still widely used as a semi-quantitative method in a clinical setting. In these steps, staining intensity is judged, as in HER2 expression, where the staining intensity is scored in a range from 0 to 3+. Even though strict reference standards are used for these, they do not reflect a linear correlation between staining intensity and protein amount [115].

1.7 The Human Protein Atlas

The Human Protein Atlas (HPA) is an online, open-access database that contains over 10 million high-resolution images of immunohistochemistry (IHC) stainings from tissue microarrays [116]. It covers 44 different normal tissues and the 20 most common cancer forms and covers 87% of human genes. Each gene is typically targeted with an in-house generated and a commercial antibody [117–119].

In the past decade, abundant gene expression data has been generated from different high-throughput systems because most of it comes from homogenized tissues or cell lines; the information on which cell type, ECM proteins, and the overall localization of the expression pattern within a tissue will typically be missing from the data.

The advantage of the human protein atlas over other databases is that the images contain information on the spatial distribution and localization of protein expression in the different tissues. Thus, the HPA can be used for a more precise localization of protein expression than other databases can provide. For example, it can be used to analyze the expression of proteins in the extracellular matrix, single-cell types, and cell subtypes in tissues.

Due to the importance of antibody specificity, validation scores for each antibody used are provided in the database [120]. Two types of validation scores can be assigned: a standard and a formally validated one. The formal validation follows stricter criteria by following the recommendations outlined by the International Working Group for Antibody Validation [121].

The standard validation score is given for every antibody. It is based on consistency with scientific literature and with RNA sequencing (RNAseq) data. The literature conformity can be supportive or not supportive and is based on whether the expression data conforms with the literature and bioinformatic predictions. Uniprot [122] is the primary source for the literature data, while more in-depth research is used when necessary. The RNAseq data is also either supportive or not and is based on internal and external RNAseq data [120].

When both RNAseq and literature/bioinformatics data are supportive, the antibody receives an overall "supported" score. When one is supportive and the other uncertain or non-supportive, the overall score is approved, and if both are uncertain or non-supportive, the overall score is labeled uncertain.

The stricter, "formal" validation score relies on two separate methods analyzed in both tissues and cell lines. It is only performed on antibodies with supportive tissue atlas data. The method is based on the correlation calculated between RNA expression and protein expression from either cell lines or tissues. In the case of the cell lines, imaging software evaluates the staining, while the tissues rely on annotated data. The "independent antibody" method uses the correlation between two in-house generated single target antibodies for validation [120].

To target such a vast number of targets, the HPA developed their own polyclonal antibodies (Atlas Antibodies Advanced Polyclonals) for primary antibody detection. They employed several steps to ensure their specificity and reproducibility. To select the epitopes, they used an antigen design software that selects a Protein Epitope Signature Tag (PrEST) sequence, ranging between 50 to

150 amino acids, which reduces the possible cross specificity to other human proteins. While full-length recombinant proteins frequently augment the risk for off-target binding, the PrEST antigen minimizes unwanted specificity by evading regions with high regional or local sequences to other proteins. They further include a unique purification process by using the recombinant PrEST-antigen as an affinity ligand to increase the reproducibility [123].

Since polyclonal antibodies may exhibit variation in immune responses between batches, they use a strict quality control procedure to ensure that both the specificity and functionality are retained between batch numbers. They are, therefore, produced through a standardized process to ensure consistency across batches. This involves analyzing each new lot in parallel with a reference lot, employing consecutive sample materials for all approved applications.

By analyzing the Human Protein Atlas, valuable information can be gained about the protein expression of molecules in the tumor stroma and the invasive edge of cancers. At the same time, careful interpretations are warranted due to the unreliability of some antibodies. The HPA provides an online website interface to explore the protein expression of various genes. However, this interface does not allow an easy method to compare images side by side, nor to do this on a large scale, for instance, to analyze a list of proteins of interest.

1.8 Digital representation of images

Digital images are made up of pixels, which can be understood as the most basic components of an image. It received its name from the combination of "*pix*" (short for pictures) and "*el*" short for element [124]. They are digitally represented as two-dimensional matrices of intensities of each pixel. These will have a certain number of rows and columns based on the resolution of an image. For example, an image with a resolution of 1920 x 1080 will have an equal number of columns and rows of pixels. Each pixel is represented by a value determined by its intensity range. A binary image will only have two values (white or black), while grayscale images will have four values in its most basic 2-bit form (white, gray, gray, black), 256 values for 8-bit images, and 2^{32} in 32-bit images.

Color images are represented by each pixel having three separate values based on the intensities in Red, Green, and Blue (RGB) channels, and an additional transparency value can also be included (RGBA). These are based on the trichromatic vision of humans. However, other approaches exist to represent color

images, such as with the HSB color spacer, in which Hue, Saturation, and Color are used [125].

The pixel values can be represented by a histogram, creating a graphical representation of the number of pixels in each intensity. The vertical axis represents the number of pixels for each intensity in the horizontal axis. For example, black pixels are on the far left of the histogram, while white is on the right. A color histogram can be made similarly by combining three histograms based on their three pixel values (RGB).

These representations can be helpful for image processing, as they allow for the adjusting of the contrast of an image, for example, histogram equalization that modifies the histogram by creating a more uniform distribution and can be used to enhance the contrast of the image [126, 127]. Histograms are also valuable for thresholding and binarization of images by helping to identify optimal thresholding values.

1.9 Image segmentation

Image segmentation is the process of partitioning parts of an image into separate objects. This can be useful in detecting tumors in CT scans or cancer cells in histopathology. The most basic method for this is the thresholding method, which divides the pixel values into two categories, creating a binary image with only two values. Several methods have been developed for this, of which the simplest is to select a threshold value manually, and values below or above are assigned the maximum or minimum values. Other methods have been developed to automatically threshold images and identify the most optimal thresholding value. One commonly used method is Otsu's thresholding [128], which systematically explores different thresholding values until it identifies the value that in a histogram would maximize that spread between the two groups. It divides and minimizes the variance inside the grouped values and maximizes them between the groups (Figure 3).

Thresholding methods work only on images with pixels of a single value, like grayscale images, represented by a single histogram. Medical histological images are often color images, having three color channels. To threshold these images for image segmentation, the colors must be separated. One established method for this was described by Ruifrok and Johnston [129] and implemented in most image processing software like ImageJ, Scikit Learn, and QuPath [130–132].

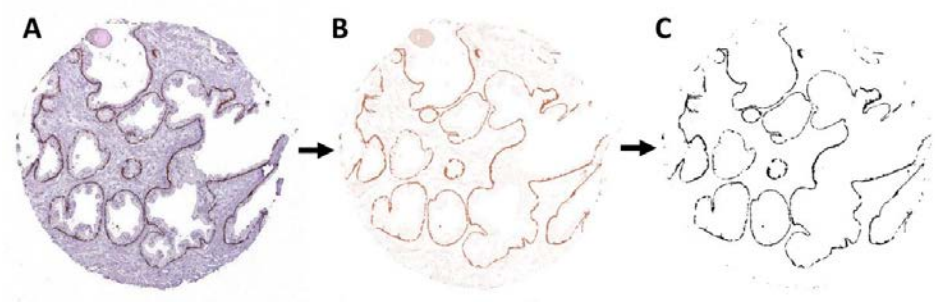

Figure 3. Image segmentation with color deconvolution and thresholding. A) IHC tissue microarray showing CK5 expression with DAB and counterstaining with Hematoxylin. B) Color deconvolution was performed to select the DAB staining; however, some background noise remained. C) Otsu's global thresholding [128] was applied to get a binary image. Image credit: Human Protein Atlas; analysis was performed with the Fiji software [133].

While there are multiple methods for image segmentation, their approaches are limited by the complexities of images. It is challenging to define features in an image or to develop mathematical models that can be applied to images that inherently contain a lot of variation. Therefore, it is challenging to generalize the methods to a large set of images [134].

1.10 Deep learning and neural networks

Deep learning, a subfield of machine learning, has successfully met the challenge of autonomously identifying features in images. This approach is grounded in computational methods that vaguely emulate the connectivity of neurons. Its history traces back to 1943 when scientists first proposed a computational model of a neuron, known as the artificial neuron [135]. The perceptron was later introduced as the first model that could learn from data [136]. However, initial models faced a substantial hurdle: a limited ability to learn from data. This challenge was later mitigated using back-propagation, which employs gradient descent to enhance the learning of the network [137, 138]. Subsequently, neural networks were shown to be successful in image recognition tasks, imitating the human visual cortex [139]. Introducing a convolutional layer further enhanced image recognition capabilities [140]. This convolution layer consists of trainable filters that help extract and learn the images' features through convolution. A pivotal moment occurred during an image classification competition, where modifications of the convolutional layer outperformed all competitors. Following this, neural networks with feature learning became the standard for image recognition and classification [134, 141]

In a typical feedforward neural network (Figure 4), for example, when recognizing a handwritten number, input, in the form of number matrices representing each pixel, is passed through the network's input layer during the training steps. With varying strengths, this layer connects to the next layer of nodes, and those nodes, in turn, connect to all nodes in the subsequent layer. Numbers are randomly passed through until they reach the output layer, where nodes correspond to expected outputs.

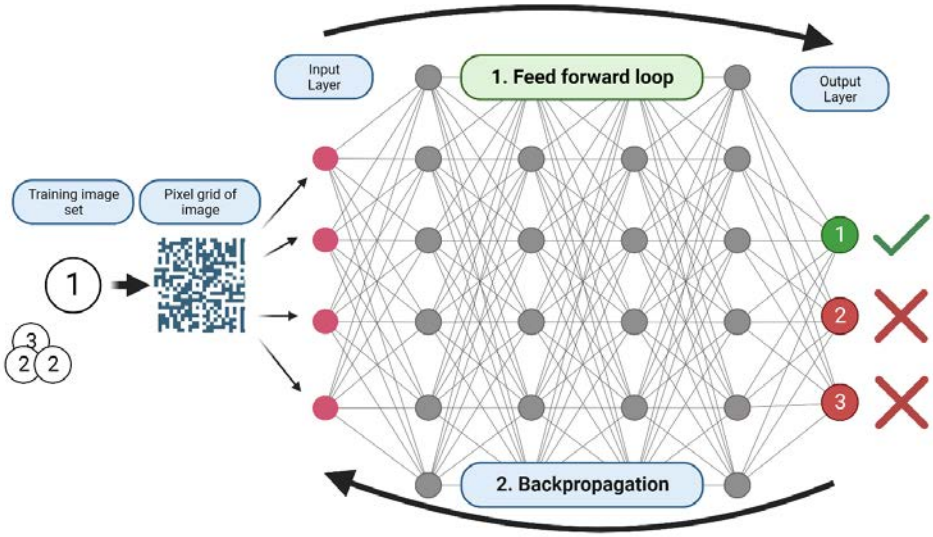

Figure 4. Schematic of a simplified neural network used for image classification. The process begins with the introduction of a training image set. Each image from this set is digitized into a matrix of numbers, where each value signifies the intensity of a corresponding pixel. This matrix feeds into the input layer. Nodes within the network, represented by circles, are interconnected by weighted links to nodes in subsequent layers. Upon completing a training epoch, the error is determined by comparing the predicted output to the actual target in the output layer. The weights of the connections are then adjusted through a process called backpropagation to minimize this error. Created with BioRender.com

Given that this is training data, the accuracy of the output layer is verifiable, enabling error calculation. This error is then backpropagated from the output layer to the input layer. The weights connecting the nodes are adjusted using gradient descent to minimize error [137, 138]. This step is reiterated for several passes through the network in "epochs." The capabilities of these basic feedforward neural networks were later amplified by appending a convolutional layer at the start of the process, thereby enhancing the extraction of image features [140]. Training these networks on large image datasets demands significant computational power. However, many of these models are openly accessible and can be retrained on alternative image sets in transfer learning, which decreases time and computational resources.

2 Research aims

This thesis aims to investigate the composition and role of the tumor microenvironment and, in turn, the effect that it can have on cancer cells. As described in the introduction, the importance of the tumor microenvironment has emerged to have a significant role in carcinogenesis, its ability to invade normal tissues and to metastasize. However, much of our knowledge stems from studies of model systems, like cell co-culture studies or animal models. Even though valuable insights have been gained from these, the model systems might not be able to fully reproduce the complexity existing in the tumor microenvironment and its interaction with cancer cells.

To complement the existing knowledge in human cancerous tissues, we set out to develop a method to analyze the composition of the TME and its function in human tissues and to investigate the presence and extent of influence it can have on cancer cells.

The specific aims of this thesis were to:

- To develop a software to enable us to analyze the Human Protein Atlas at a large scale (**Paper IV**).
- To identify new proteins expressed in cancer associated fibroblasts in the tumor microenvironment, compare their expression in different tumor types, and infer a functional role from these protein expressions (**Paper I**)
- To investigate the connection of RhoA expression in cancer associated fibroblasts in model systems by creating RhoA knockout fibroblasts and to study their effects on cancer cells *in vitro* in co-culture assays and *in vivo* animal models (**Paper III**).
- To analyze the expression of the proteoglycan Decorin in the tumor microenvironment and compare its expression to other closely related proteoglycans (**Paper II**).
- To analyze a set of pancreatic ductal carcinomas that invade the small intestine where phenotypic plasticity can be observed to explore the extent to which the ability of the normal microenvironment effects of cancer cells can be observed outside of model systems (**Paper IV**).
- To develop a new method for the large-scale analysis of the HPA using deep learning to enable high throughput analysis of the HPA and to use the method to identify new markers of prostate basal cells (**Paper VI**).

3 Materials and methods

3.1 Ethical considerations

Our research involved analyzing tissue samples from healthy individuals and cancer patients. We also conducted animal experiments to analyze the effect of our findings *in vivo*. All necessary ethical approvals were obtained before beginning any of the studies.

Our research relied heavily on the Human Protein Atlas online database (**Papers I, II, V, and VI**). All the used human tissue samples were collected and handled according to Swedish laws and regulations. The tissue samples were collected at the Pathology departments in Uppsala University Hospital as part of the Uppsala Biobank. Further, all the tissue samples were anonymized. The Uppsala Ethical Review Board approved the study.

Because the database creators received the proper ethical permissions from their local review board, the sample collection followed Swedish law, and the data was anonymized, I feel confident that using the data obtained from the database is ethical. In addition, we also developed two methods that can improve the functionality and expand the use of this database (**Papers V and VI**). These tools can enable other researchers to take advantage of the database and make its use even more widespread.

Paper IV used human pancreatic cancer tissues surgically excised during routine Whipple resections. In addition, we also acquired survival data for these patients from electronic health records. The research did not result in additional patient interventions, and the data was anonymized. Since the diagnostic and therapeutic procedures were standard of care, they did not require special patient permission. Written consent was also considered unfeasible as the retrospective study went back to 2002, and the diagnosis has a poor prognosis. The samples used for histopathology only included those where patients had consented to the inclusion and storage into the local biobank.

In addition, we also analyzed the presence of the phenotype switch that we observed in human cancers in animal models. In total, we looked at six mice previously used in another research study, so no additional harm had come to these animals with our analysis. However, we could use their sacrifice for additional insight into cancer research.

In **paper III**, animal models we used animal models to delve deeper into our findings from **paper I**. Initially, we used cell culture studies to assess the implications of knocking out RhoA. Animal studies inherently lead to harming animals and, therefore, demand meticulous planning and strict ethical adherence. Therefore, careful considerations are necessary on the value and necessity of the research. Due to the intrinsic complexities of cancer biology, particularly the complex interactions between tumors and their microenvironment, we considered using *in vivo* models important to fully understand the effects in living organisms, thus justifying the use of animals. As it leads to the harm of animals, it is of great importance to adhere to high animal welfare standards, limit the animals' suffering as much as possible, and use strict scientific rigor to ensure that the potential scientific gains are of good quality. I am confident that we could adhere to these principles.

3.2 Paper V. AtlasGrabber: a software facilitating the high throughput analysis of the human protein atlas online database

To allow us to perform a larger-scale analysis of the HPA database and exploit its extensive capacity, we developed a Windows desktop application, AtlasGrabber, to enable this.

We wrote the code in C# coding language and released it on GitHub [142] under Gnu Public license v3 [143]. Its primary function is to streamline the analysis of IHC-stained tissue samples from the HPA, focusing on a specified gene list. Based on a provided gene list, it can systematically present images side-by-side, thereby facilitating the categorization and sorting of proteins of interest into different lists that can be saved and later retrieved. It, therefore, allows concurrent analysis of a predefined list of proteins across up to four distinct tissues in both tumor and normal tissue samples, thereby enabling comparisons between different tissue stainings for a single antibody. In addition, we added a feature to analyze the entire database from HPA, available only as a downloadable XML file [144]. This tool, the XML parser, can find the complete list of image links for a particular tissue and includes the names of genes and antibodies for any of the tissues included in the HPA.

To use the AtlasGrabber, a text file (.txt) containing a list of Ensemble IDs for the intended gene analysis is required. These lists are generated by extracting the names with the XML parser for a complete list or via keyword searches within the

HPA search function, followed by file exportation. A comprehensive guide, with detailed video guides, can be located on the GitHub page's Readme file [142].

The AtlasGrabber can be accessed on its GitHub page [142] in an executable file that can be directly run without the need for installation, or by using the source code, it is possible to compile it. It is tested to work on Windows 8, 10, and 11. Optimal usage is achieved with a large screen, high-definition display, as it will maximize the screen area usage by dynamically recalculating each window's occupied area.

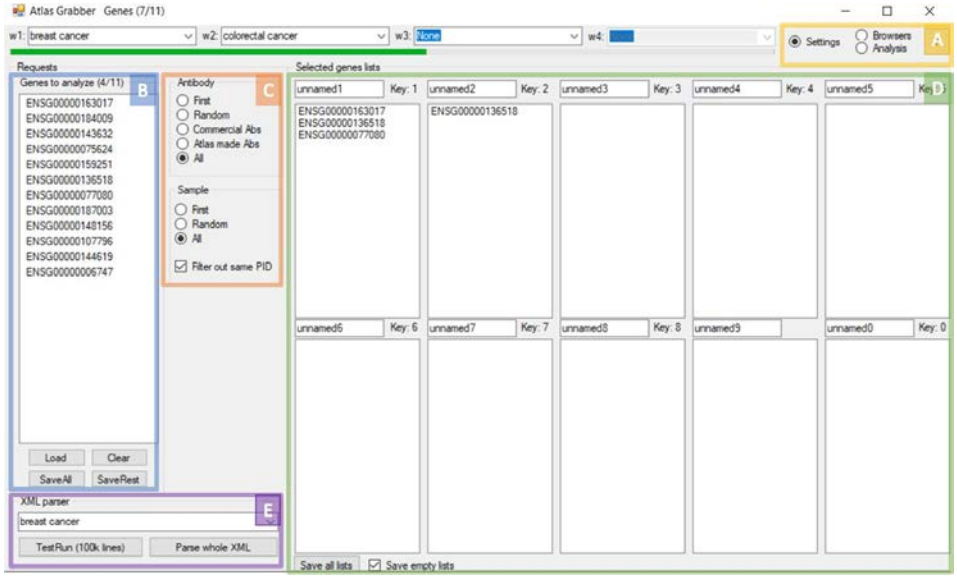

Figure 5. The AtlasGrabber with the "Settings" open. A) Windows are selected at the top right corner. B) Displays the loaded protein list. C) Additional filters can be selected depending on antibody type and sample. D) Selected proteins are saved to these lists that can be named and saved. E) The XML parser feature extracts complete protein lists from normal or tumor tissues.

The application has three windows: analysis, settings, and browser (Figure 5). On startup, it launches with the "Settings" window, wherein users upload their text file containing the gene list of interest. Here, added settings include viewing image slides of all antibodies or viewing only commercial or HPA-produced in-house ones. It can also filter the images to view only one image per patient sample (as typically, from each patient, two images are provided). In addition, the settings window is used to view the gene lists for storing selected genes. A keyboard shortcut is assigned to each list from 0 to 9. Pressing that key during the analysis will save the current protein to that list. When saved, the list is stored in the same folder as the executable, and the newly analyzed gene names are appended to existing ones, allowing users to continue from a previous analysis.

To initiate the image loading and viewing process, the "Analysis" window is used. Here, images related to the protein from the initial list are shown, and one can scroll through each image for every antibody targeting the protein. We assigned designated keyboard shortcuts to navigate through images, antibodies, and proteins, though the mouse scrolling wheel can also serve this purpose. In the "Settings" window, one can view the protein under analysis and view gene IDs allocated to various lists.

In the "Browser" window, users can view the HPA website related to a specific antibody in a web browser, facilitating access to summaries about the gene or antibody. If an antibody is of interest during analysis, rapid access to relevant information from the HPA, such as antibody type, antibody validation, and descriptive summaries of the protein can be viewed. A progress bar and the current protein analyzed can be viewed at the top of the application window. A "Help" link directs users to the GitHub page's Readme file, offering detailed instructions and tutorial videos.

In the "Settings" window, the XML parser can be accessed. This enables the extraction of data subsets (both normal and tumor tissues) and is saved in a .csv file.

To illustrate the application's utility, we aimed to identify new biomarkers for basal cells of the prostate gland. These cells encircle the normal glands in the prostate but typically disappear in prostate cancer as the cancer cells invade and break through this cell layer. Pathologists commonly use three IHC markers—CK14, CK5, and P63—to identify prostate basal cells in diagnosing prostate cancer, as their absence indicates invasiveness [145–148].

With the XML parser function, we accessed the protein Ensemble IDs [149] for normal prostate tissue, selecting a subset from the gene list for analysis. The list was analyzed using the AtlasGrabber's "Analysis" function in normal prostate tissue. The proteins that showed a positive staining for prostate basal cells were selected. In a comparison analysis, we then compared their expression in prostate cancer.

3.3 Paper I. Novel signatures of cancer associated fibroblasts

In a preceding study in our research group, we identified 1,033 genes from gene expression analysis of two pairs of isogenic, inhibitory, and non-inhibitory fibroblasts regarding their effect on cancer growth [150]. One pair comprised the

telomerase immortalized BjhTERT fibroblast cell line [151], while the other pair was from donor fibroblasts. Both pairs exhibited inhibitory and non-inhibitory characteristics established from fibroblast-cancer cell co-culture studies. Upregulated genes were selected from these inhibitory and non-inhibitory fibroblasts.

From these 1,033 selected genes, 759 were selected that were identified in the UniProt Accession numbers [152]. To expand this protein list with relevant proteins, first neighbor interactors were selected from the BioGRID and Human Protein Reference Databases [153, 154], thus ending up with an additional 1,892 proteins identified as possible interactors with the original list of proteins.

Since these differentially expressed genes were proposed to play a role in fibroblast function in the TME, we set out to analyze their expression in cancer associated fibroblasts by comparing normal and cancerous tissue stroma in the Human Protein Atlas online database. To enable this, we used an early version of the AtlasGrabber software [155]. We initially analyzed the protein expression of these genes in basal cell carcinoma compared to normal skin tissue. Skin and basal cell cancer were selected, as the skin fibroblasts can be easily visualized, and basal cell cancer typically forms clear cancer island nests with clear boundaries to the stroma. Proteins that expressed a clear and strong staining in the basal cell carcinoma stroma but not in the normal skin were selected. An additional selection was made by the number of samples that expressed CAFs.

To investigate how these genes are expressed in CAFs in other cancers and to what extent, we analyzed their expression in the stroma of squamous cell carcinoma, breast cancer, colorectal cancer, and lung cancer. These cancers were also selected based on the ability to identify the tumor stroma clearly. As CAFs have a similar phenotype to myofibroblasts, we investigated how our newly found CAF markers compared to normal tissue fibroblasts that included fibroblasts with a myofibroblastic phenotype, including bone marrow fibroblasts, mesangial and peritubular fibroblasts of the kidney, subepithelial fibroblasts of the small intestine, and intra, and interlobar fibroblasts of the breast.

3.4 Paper II. Decreased decorin expression in the tumor microenvironment

Our study on CAFs revealed that an ECM protein, Decorin, was highly expressed in normal tissues but seemingly absent in the basal cell carcinoma tumor stroma, opposite to our CAF markers. It was striking that it seemed nearly absent in the

tumor stroma we had seen. The research on Decorin indicated a significant role in carcinogenesis and a clear tumor inhibitory role [54]. We were therefore interested in looking at its expression in a wide range of tumor types. For this, we could again take advantage of the HPA and our software tool, the AtlasGrabber [155]. The antibody for Decorin expression in the HPA had supportive validation.

To improve the objectivity of this analysis, we developed an additional in-house tool, the Protein Expression Quantifier, to objectively quantify the protein expression in the IHC-stained images. It analyzes the DAB IHC signal from images by image inversion and automatic thresholding, after which the DAB expression is quantified and charted.

The striking absence of Decorin in the tumor microenvironment led us to look at additional similar proteoglycans of Decorin: Asporin, Biglycan, and Osteoglycin. We analyzed the expression of these in skin and breast cancerous and normal tissues.

3.5 Paper III. RhoA knockout fibroblasts lose tumor-inhibitory capacity in vitro and promote tumor growth in vivo

In **paper I,** we identified 12 new markers of cancer associated fibroblasts. We found that four of the 12 identified markers were related to Rho-signaling. Therefore, we set out to investigate the closely related role of RhoA in fibroblasts and its effects on cancer cells in both *in vitro* and *in vivo* model systems.

We established a RhoA knockout (KO) immortalized BjhTERT fibroblasts [151] line with the CRISPR/Cas9-lentivirus knockout method [156]. We used a mixture of three vectors, RhoA1, RhoA2, and RhoA3, to transduce the cells. The control group was made with fibroblasts infected with empty vector lentiviral particles. The protein level was evaluated using Western blotting. For cancer cells, the prostate cancer cell line PC3 was used [157].

Using a tumor inhibitory assay, we studied the effect of the RhoA knockout fibroblast on tumor growth. For this, we plated the fibroblasts and cultured them for five days until a confluent monolayer was formed. We then plated 200 fluorescently labeled PC3 cells on the monolayer. The control group was without the fibroblast monolayer. The analysis was performed with an earlier established automatic immunofluorescence microscope system, measuring at a single cell level [158].

We used the female SCID mice models to investigate the effect of the RhoA KO fibroblasts on cancer growth in vivo. We injected them with PC3 with a non-tumorigenic dose of PC3 cells (20,000) alone, together with the control wild type (WT) RhoA fibroblasts, and finally with RhoA KO fibroblasts. The occurrence of tumors was analyzed up to 80 days post-injection. The tumors were measured with a caliper to mm^3 size.

We used a live-cell motility assay to measure the effect on cancer motility. KO and WT RhoA fibroblasts were cultured in six-well plates. H2AmRFP-labeled PC3 cells were then added and co-cultured. Total Infernal Reflection Fluorescence (TIRF) Microscopy recorded time laps of cancer cell motility for 65 hours.

To study the effect on cancer cell growth, we co-cultured the RhoA KO and WT fibroblasts in a 3D collagen matrix growth assay with PC3 cells. A clustering Index was calculated to quantify the growth of the tumor spheres.

To further investigate the effects of these knockout cells, we studied their secretory function when co-cultured with PC3 cells. We added labeled PC3 cells to fibroblast monolayers and sorted the cells with fluorescence-activated cell sorting (FACS), after which we isolated the RNA and analyzed it with the Affymetrix Microarray. Differential gene expressions of the RhoA KO and WT fibroblasts before and after PC3 confrontation were measured. PC3 cells were cultured with the RhoA KO fibroblasts, WT fibroblasts, and both with and without the PC3 cells. The results were validated with qPCR analysis.

Finally, to investigate the cytoskeletal function of the RhoA knockout cells, we measured the stiffness and contractile force of these fibroblasts using atomic force and traction force microscopy.

3.6 Paper VI. Identification of novel protein markers of prostate basal cells by application of deep learning to images from the Human Protein Atlas

We developed a new method based on deep learning to enhance our ability to analyze protein expression in the HPA on an even larger scale. Using the AtlasGrabber's [155] XML parser feature, we accessed all the available images for normal prostate tissue from the HPA database. Out of these, we manually identified 70 images that showed positive staining for the prostate basal cells. Additional images were selected that represented the non-basal cell staining. These formed our two training datasets.

As the TMA images vary in color intensity and their tissue architecture, we performed image segmentation to isolate the DAB staining of the images. We used those images for the training and classifier. This image processing was done using the Fiji software [133], a distribution of ImageJ2 [130].

First, we applied background subtraction with color correction, followed by color deconvolution based on manually identified color vectors from regions of interest from representative images [129, 159]. Due to the varying color intensity, two separate color vectors were used. Once color deconvolution was applied, we used binarization of the images with the Otsu [128] and Minium methods [160] since we noticed that after using different color convolution methods, no uniform binarization fit all images uniformly. The image processing was applied to all images. Thus, we ended up with four new binarized images for each original image. We made an additional manual selection of the images for the training set that were selected based on which showed the clearest staining patterns. Finally, we ended up with 164 images for the training set that showed a basal-like staining pattern and 500 images that showed a non-basal-like staining, which formed our two classes. The discovery analysis was performed on a total of 185 850 images, of which 39 104 made up the testing sets used to evaluate the performance.

We used GPU-accelerated TensorFlow for the classification and tested two neural networks: MobileNet [161] and Inception V3 [162]. Transfer learning was used to retrain the networks on our training image datasets. The images were classified, and a score of either basal-like or non-basal-like was assigned to each image. From the first training set results, we selected 53 misclassified images and retrained the network with these images (Figure 6).

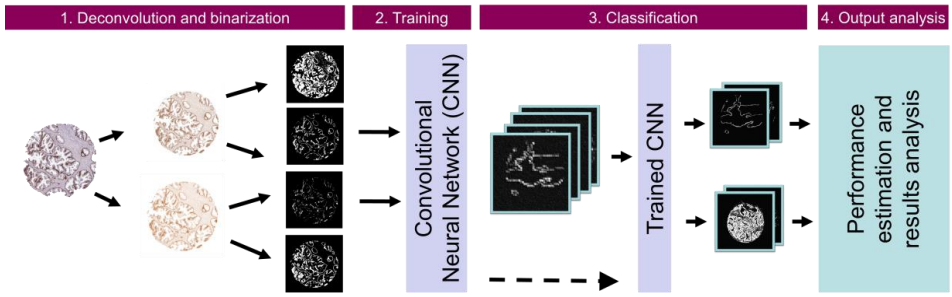

Figure 6. The method of image processing and classification. 1) Deconvolution and binarization were applied to all images. 2) The CNN was trained with transfer learning. 3) The trained classifier was used to classify all the images as either basal-like or not. 4) Additional manual selection was done on the classified images, and performance was measured using the testing set.

At the cutoff of 0,95 prediction for basal cell phenotype, the network identified 3,053 images out of the original 185,850 images. Upon inspection of the results, the classifier had made some errors, like not being able to easily distinguish between basal cell type and glandular epithelial cells, as both showed a similar staining pattern after our image processing steps. We finally made a manual selection of the images and selected 53 proteins that showed the strongest and clearest staining.

Out of those, several were based on the same original image. Each antibody could have several corresponding images that showed a similar staining pattern, and each protein might have several antibodies targeting it that also showed a similar staining pattern.

To investigate whether our identified genes would overlap with RNA expression data, we compared our results to an existing gene expression dataset that we published in a previous study [163]. This dataset was based on single-cell RNA expression (scRNAseq).

We also performed a network enrichment analysis (NEA) with FunCoup 3 [164] to see if we could infer a functional role from our found markers. For this, the functional gene sets were from Gene Ontology (GO) [165, 166], Disease Ontology (DO) [167] and Reactome Pathways, and the network was based on Pathway Commons v9 [168, 169].

Finally, we investigated whether the basal cell markers we identified were present in prostate cancer. For this, we used the AtlasGrabber software [155] which allows for the easy comparison of protein expressions in different tissues in the Human Protein Atlas database. A staining score was assigned based on the low, medium, or high for all the antibodies for each protein, based on a subjective interpretation of the strength and quality of the stainings.

3.7 Paper IV. Stabilization of the classical phenotype upon integration of pancreatic cancer cells into the duodenal epithelium

In vitro model systems have indicated that cancerous cells can change their phenotype and even normalize depending on their tissue context [80] However, it is unclear to what extent this can occur *in vivo*. An observation had been made by pathologists at the pathology clinic at the Karolinska University Hospital, where they had examples of certain pancreatic cancers that are typically locally very invasive, where the pancreatic ductal adenocarcinoma (PDAC) cells invaded the

small intestine. Seen with multiplex immunohistochemical stainings, they integrated into the intestinal crypts and villi. Here, their immunostainings gradually changed, and they acquired an immuno-phenotype of more indolent, classical PDAC and that of the normal intestinal cells.

To further study this and to quantify the extent to which this can be observed, we performed a retrospective search for cases that showed this phenomenon. Cases were identified between 2008 and 2020 where PDAC cells infiltrated the small intestine and showed a change in their IHC staining patterns.

For these cases, serial multiplex immunohistochemistry [170] and quantification were performed separately in the mucosa and submucosa of the intestine to differentiate the IHC staining patterns.

In addition, FFPE sections from a cohort of pancreatic cancer mouse models [171, 172] were analyzed, and one animal was found to have a similar intestinal invasion as described in our patient material. With IHC, we analyzed the expression of the high-motility group AT-hook 2 (HMGA2) protein by immunohistochemistry that was shown to be downregulated in classical PDAC [172].

4 Results and discussion

4.1 The AtlasGrabber software enables the large-scale analysis of the Human Protein Atlas database (Paper V)

To enable our research aims and to be able to analyze the HPA fast and on a large scale, we set out to develop the software AtlasGrabber. We had already used an earlier version in **paper I,** and the software went through several cycles of improvements until we made the final application. We made it available for other researchers using an open-source license [143], allowing unrestricted use and modification in research.

Previously, another research group had identified a similar need for such an application, the HPASubC application [173]. However, we had tested it prior to designing our software. We found it challenging to use and run since it only runs on Linux kernel and requires several Python scripts and dependencies, many of which are outdated. Through our extensive use of the HPA, we were also able to identify several uses in our research that it did not enable but which we implemented. For instance, the AtlasGrabber allows us to compare up to four images side by side, allows us to sort the proteins of interest into different lists, and has options to filter which images for each protein one views. It also fetches the images directly from the HPA database and only loads them temporarily; thus, it does not need to download large image sets.

We also made it user-friendly so that it is easy to run on Windows without any installation required and added easy tutorials online. In addition, we also added the XML parser functionality that allowed us to extract entire protein, antibody, and image addresses from the XML Human Protein Atlas database file, which no other software could easily do due to its large size. We used this function in **paper VI** to access all the images for normal prostate tissue.

As proof of concept, we set out to identify protein markers of prostate basal cells. We could easily identify six new basal cell markers from a limited number of genes using the software (Figure 7), demonstrating its utility. We then used the comparison function of the AtlasGrabber to compare our findings to cancerous tissues. This showed that, as expected, most of the basal cell markers were absent from prostate cancer cells, except for EMC8.

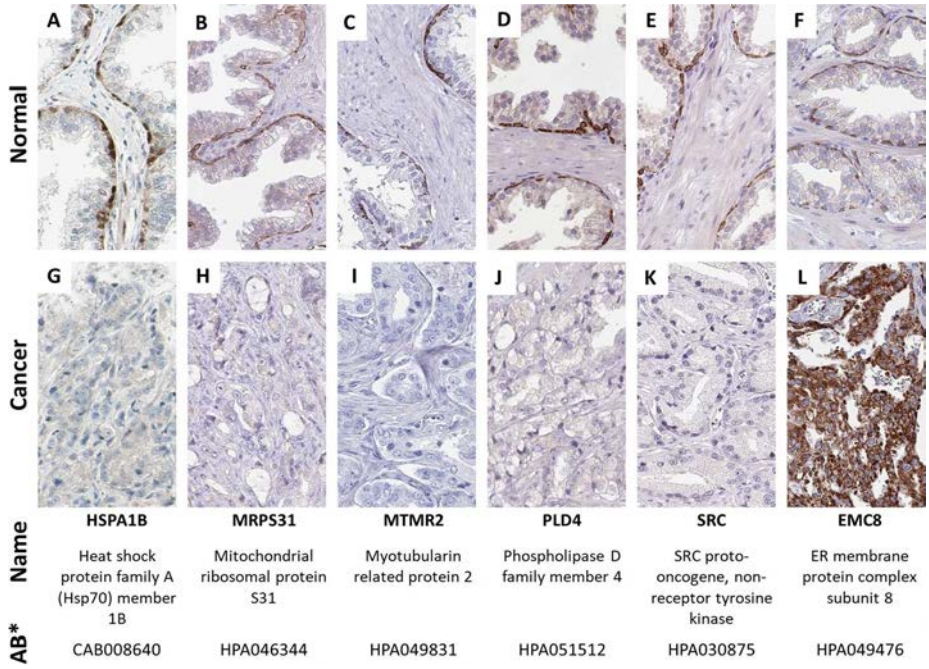

Figure 7. Identification of Novel IHC Markers for Prostate Basal Cells using AtlasGrabber Software. Images A–F depict markers with distinct and specific staining for basal cells in normal prostate tissue. Images G–L show their expression in corresponding prostate cancer, all absent apart from EMC8. The bottom section provides the name of each protein alongside its corresponding antibody ID. Image source: Human Protein Atlas.

The software enabled us to speed up our analysis, which would not have been feasible without the software. Nevertheless, a limitation of this software is the need for the user to view each image one by one and make a subjective interpretation of each of them. The HPA contains over 46 thousand images for normal prostate tissue and can have over 200 thousand for each cancerous tissue. It is, therefore, not feasible to look through all these images, even with software like the AtlasGrabber that significantly speeds it up.

This led us to explore various computer vision methods to automate this so that from selecting a few images, one could automatically identify similar images in the HPA. For this, we first explored using perceptual hash algorithms [174] that create a distinct fingerprint of an image that can be compared to each other for similarity using a Hamming distance, a technique used in image search engines to find copyrighted materials online. However, we could not get these to work well on the IHC tissue images from the HPA. However, at the time, a new technique emerged, the use of convolution neural networks that could be used to classify images. We turned to this technique and developed it in **paper VI**.

4.2 New markers of cancer associated fibroblasts (Paper I)

In **paper I**, we used the AtlasGrabber to identify new proteins expressed in cancer associated fibroblasts of the tumor microenvironment. We started with a list of 1,876 genes, and out of these, we identified twelve proteins expressed in the tumor stroma of basal cell cancer but not in normal skin fibroblasts.

The comparison in different tumor types revealed that the expression of these CAFs was not uniform in different tumor types but varied in their expression (Figure 8 and Figure 9). For example, ARHGAP26 was positive in most basal cell cancers but not equally in squamous cell cancers.

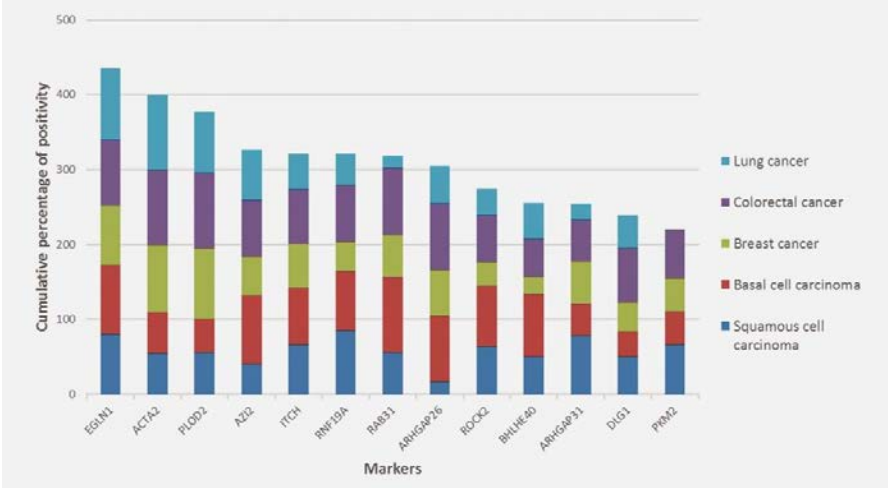

Figure 8. Each newly identified CAF marker and its expression in different cancers. For each cancer, we looked at all the available patient samples for each tumor and calculated the percentage of positive stains. We also included Smooth muscle alpha-actin (ACTA2), the standard CAF marker for comparison.

We further compared the expression of these CAFs to other normal tissue fibroblasts. This revealed that they were not expressed in most normal fibroblasts of the breast and of peritubular fibroblasts of the kidney but were expressed in the small intestinal subepithelial fibroblasts and somewhat expressed in bone marrow fibroblasts and the mesangial fibroblasts of the kidney. This could be explained by the more myofibroblastic phenotype of these normal tissue fibroblasts. A notable exception was ROCK2, which was only expressed in CAFs and absent in all the normal fibroblasts.

Exploring the function of the 12 proteins, we found that four of the genes, ARHGAP26, ARHGAP31, DLG, and ROCK2, were linked to Rho kinase signaling. This pathway has been described to have a role in cell migration, cell adhesion, and

actin cytoskeleton regulation and could, therefore, be proposed to be connected to a myofibroblastic and CAF-like phenotype [175].

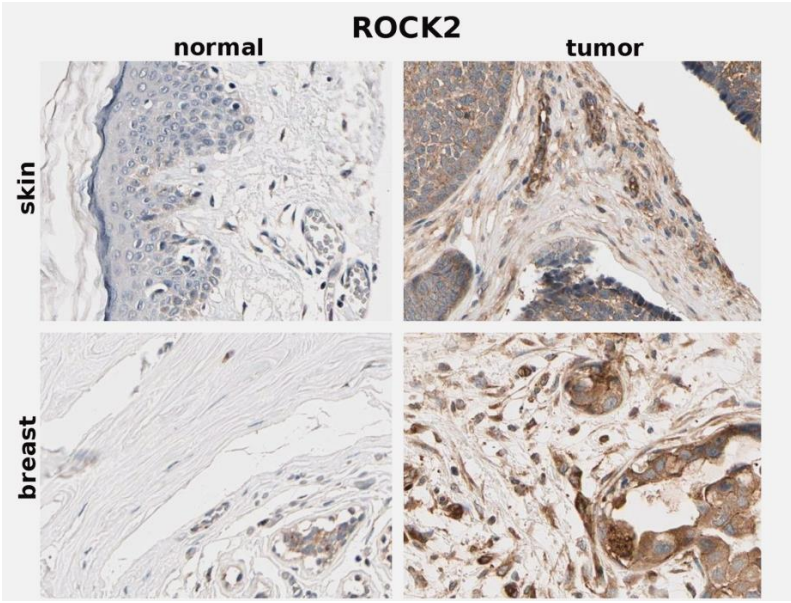

Figure 9. The newly identified CAF marker ROCK2 with its expression compared in fibroblasts of normal skin and breast in which they are not expressed, compared to CAFs in basal cell cancer and breast cancer in which they are expressed.

Performing a protein interaction network analysis, we identified a highly connected protein network, and functional analysis revealed that highlighted processes in the network included Rho- and GTPase-related processes and cytoskeleton and organelle organization.

In summary, starting from the list of differentially expressed genes from subtypes of fibroblasts that had either an inhibitory or non-inhibitory effect on cancer cell growth, we were able to use our software to identify new CAF markers which showed a clear connection to Rho kinase signaling. Limitations include that we only analyzed a subset of all proteins available in the HPA, and that the interpretations of the stainings were subjective.

4.3 Decorin expression is decreased in the tumor microenvironment (Paper II)

In **paper II**, we explored the expression of the proteoglycan Decorin that we had observed during the analysis in **paper I.** We found it expressed in normal skin ECM but almost absent in skin cancer.

We analyzed Decorin expression in normal and cancerous tissues. We found that Decorin was expressed strongly in the stroma of all the normal tissues analyzed but was markedly decreased or absent in all the corresponding tumors, although there was some variation in the expression between patient samples (Figure 10). Using our quantification method, we could objectively measure the DAB expression in the images to confirm the observations.

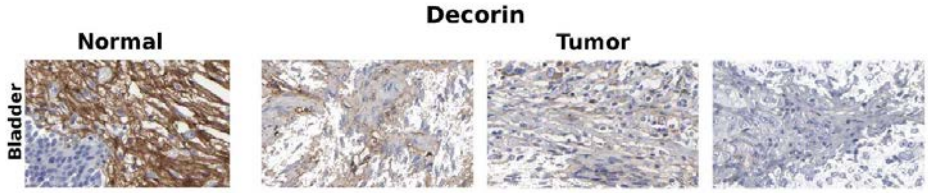

Figure 10. Decorin expression with IHC compared in normal bladder and urinary bladder cancer. A strong positivity can be seen in the normal connective tissue, while it is nearly absent in the corresponding cancer tissues.

As decorin is deposited in the ECM, mainly by fibroblasts, we postulated that one reason for the reduced expression in tumors could be that a newly formed tumor stroma could generally have a lower proteoglycan expression as it has not had time to be deposited. Therefore, we also studied closely related proteoglycans to Decorin: Biglycan, Asporin, and Osteoglycin in breast and skin tissues, and their corresponding tumors, ductal breast cancer, squamous cell cancer, and basal cell cancer of the skin. We found that Osteoglycin showed a similar expression pattern to Decorin, upregulated in the normal tissues but downregulated in the corresponding cancers. In contrast, Biglycan and Asporin showed the opposite, upregulated in the corresponding tumors. Indicating that tumor stroma can have abundant proteoglycan expression, and the decreased Decorin expression we observed might not be explained by the delay in its deposition in a newly formed tumor stroma.

The HPA, combined with the AtlasGrabber, proved to be a helpful tool for analyzing protein expression in normal and tumor tissues. It is especially valuable to analyze protein expression in the ECM as this would not be feasible by methods that look at cellular mRNA expression.

Although previous studies had shown that Decorin is decreased in the stroma in some cancers [58, 59], we added to this observation by looking at more tissues. We showed that this analysis can be easily done using our method. Our findings align with the current understating of Decorin's role in the tumor microenvironment as the "guardian from the matrix" [54] in which its absence

could promote tumor growth and invasion. The method can further analyze additional proteoglycan expression in the TME.

4.4 RhoA knockout fibroblasts lose their tumor-inhibitory capacity (Paper III)

We established RhoA knockout fibroblasts using the CRISP/Cas9 knockout model of the BjhTERT fibroblast cell line [151, 156]. Loss of RhoA was confirmed using quantitative real-time PCR and western blotting.

In our co-culture assay, we measured the effect of the knockout fibroblasts on the proliferation of PC3 cancer cells. Co-culture assays with normal fibroblasts typically decrease cancer growth [150] and we observed this with our RhoA-expressing fibroblasts. However, the RhoA knockout fibroblast had a significantly higher cancer cell growth than the control fibroblasts.

To study the effect of the RhoA KO fibroblasts on tumor growth *in vivo,* we injected SCID mice with PC3 cancer cells alone and in combination with the WT RhoA (10 mice) and KO fibroblasts (5 mice). Injecting the animals with PC3 cells alone resulted in no tumor formation, while infecting with WT fibroblast resulted in one small tumor-forming. However, our KO fibroblasts showed tumor growth in 5 out of 5 animals, with a very rapid expansion after six weeks (Figure 11).

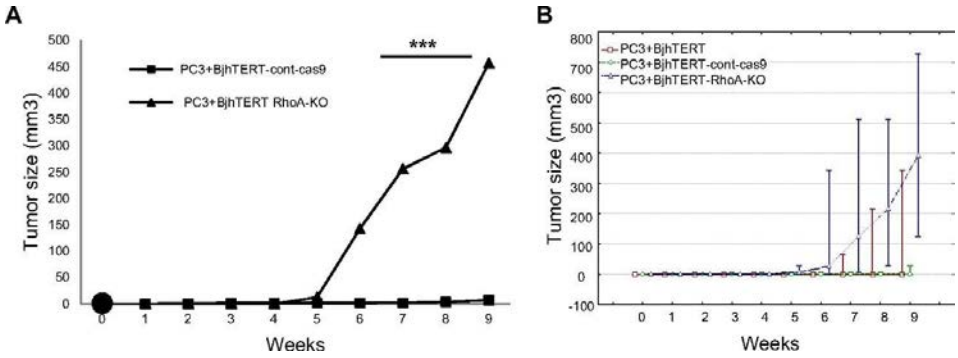

Figure 11. Tumor growth in the SCID mice injected with prostate cancer cells with either the RhoA KO or control WT fibroblasts. Cancer cells alone formed no tumors, the WT control fibroblasts only formed one small tumor, and the KO fibroblasts formed tumors in all mice.

We used a motility and proliferation assay to study the effect on the motility of cancer cells of our RhoA KO fibroblasts. We found that the effect on cancer motility was increased when PC3 cancer was co-cultured with the RhoA KO fibroblasts and showed that the cancer cells formed larger colonies (Figure 12).

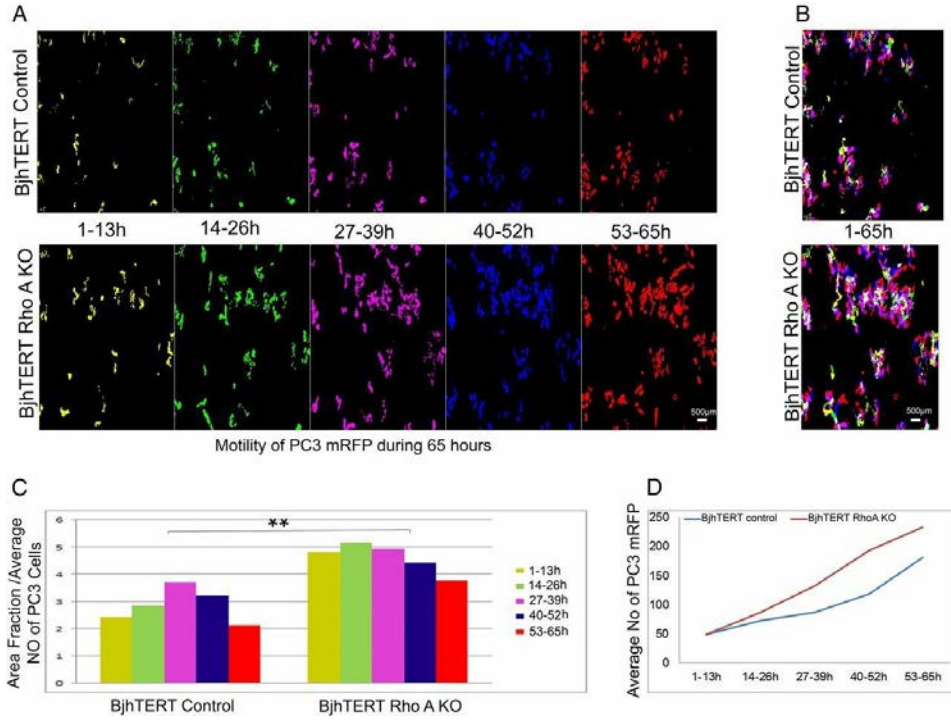

Figure 12. A) The motility assay comparing the RhoA KO fibroblasts to WT fibroblasts showed increased cancer motility with the KO fibroblasts at different time points. B) Max projection of all time points over the 65 hours. C) Kinetics of tumor-cell motility shown by calculating the areas of cell trajectories. D) Average (mean) number of cancer cells proliferating during each time interval.

In studying the change in cytoskeletal architecture and its effects on cellular contractile forces and stiffness, we observed that the RhoA KO fibroblasts showed less regularly shaped cells compared to WT fibroblasts. They also showed fewer actin stress fibers, fewer focal adhesions, and reduced α-SMA expression.

With traction force and atomic force microscopy, we measured the mechanical properties of the RhoA knockout fibroblasts. We showed that the KO fibroblasts had significantly reduced contractile forces. At the same time, the cell stiffness was more homogenous and evenly distributed in the KO fibroblasts compared to the control fibroblasts, which also appeared significantly stiffer.

Using gene expression analysis of the fibroblasts KO and WT, co-cultured with PC3 cells, we could further show that the KO fibroblasts showed an increase in proinflammatory gene signatures upon cancer cell contact.

In a 3D coculture assay, we found that KO fibroblasts resulted in the PC3 cancer cell line having a higher clustering index, meaning they were more compact than in the control group, which showed that cancer cells were more dispersed.

In **paper I,** we identified several CAF markers connected to Rho signaling. In summary, in these studies, we found that knocking out RhoA showed that the fibroblasts decreased their inhibitory capacity on cancer cell growth *in vitro* and *in vivo*. This indicates a significant role in RhoA's tumor inhibitory role.

Our observations of its expression in the CAFs in the tumor microenvironment could infer a process to inhibit tumor progression as a tissue response to malignancy. This aligns with observations in the metastasis of colorectal cancer to the liver. These observations grouped liver metastasis into two distinct types: ones accompanied by a strong desmoplastic rim - a peri-metastatic capsule, and those without. Survival data from these patients indicated a better prognosis for patients with metastasis with a strong desmoplastic reaction, indicating that it had a potential inhibitory and protective role in those patients. In addition, they could also show that this rim originated from the liver as a response to injury caused by cancer metastasis [176–178].

4.5 Deep learning enables the high throughput analysis of the HPA and identifies multiple new markers of prostate basal cells (Paper VI)

We tested the MobileNet [161] and Inception [162] networks and found a slightly better performance with the MobiNet network. Both 250 and 4000 training steps were used, but we found only a slight improvement with less than 0.03 difference with the increased training steps measured by AUC.

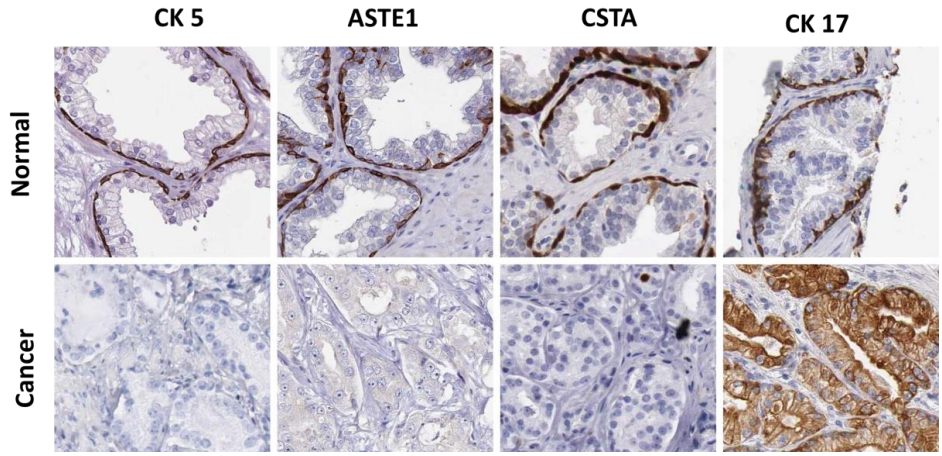

Figure 13. Examples of prostate basal markers in normal and cancerous tissues identified by a deep learning classifier. The bottom row shows the comparison to prostate cancer, where they are mostly absent.

With a cutoff of 0.95, we could classify the 185 850 images, and after a retraining step, we ended up with 3053 images classified to the basal-like phenotype. With

a final manual curation, we selected 53 markers specific to basal cells. Amongst the identified proteins were the commonly used markers in routine pathology, CK14, CK5, and P63 [145–148]. Forty-four of these markers had not previously been described as IHC markers of prostate basal cells (Figure 13).

Our gene enrichment analysis showed significant enrichment for biological functions like epithelial keratinization, development, and regulation.

Comparing our data to existing scRNAseq data on prostate basal cells [163] revealed that only 9 of our identified markers overlapped. Comparison to prostate cancer showed that most of our markers were, as expected, absent in cancer, with a few exceptions to this (Figure 14).

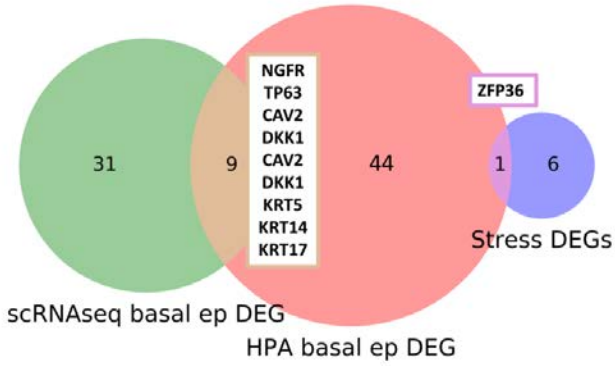

Figure 14. Venn diagram comparing our identified markers of prostate basal cells compared to results from scRNAseq data. While nine proteins overlapped and were identified by both methods, most did not.

In summary, our method shows that based on a limited number of images, it is possible to identify similar images in the protein atlas and identify multiple new IHC markers for that specific cell type. Comparison to gene expression data showed we could identify similarly expressed genes; however, the majority differed. This shows that both methods are complementary to each other.

Reasons for the differences could be explained by the known differences between gene and protein expression, which can be explained by post-transcriptional mechanisms, such as RNA splicing, editing, and mRNA stability, as well as variations in translation efficiency, protein degradation rates, and cell-specific factors [179–183]. Limitations of antibody specificity intrinsic to affinity-based methods such as IHC could also explain the difference. We could observe this as different antibodies used the HPA did not always show the same stainings for the same protein.

4.6 The normal tissue environment can influence cancer cells to change their phenotype (Paper IV)

We identified 20 patients between 2008 and 2020 who had undergone Whipple resection at the Karolinska University Hospital and had confirmed PDAC. At the time of data collection, median survival was 559 days after surgery, and only one patient was still alive at the time of data collection.

In these cases, cancer cells had infiltrated the duodenal epithelium to varying degrees and had integrated into the epithelial layer of the normal small intestine. Notably, the normal mucosal architecture of the intestine remained intact in all cases, with only mild reactive atypia. To confirm the pancreatic origin of these cells, we utilized immunohistochemical markers such as SMAD4 loss and P53 positivity [184].

Notably, the cancer cells that integrated into the intestinal mucosa underwent a phenotypic transformation. They exhibited well-differentiated features and, to some extent, regained their polarization, in contrast to the submucosal cancer cells. Typically, pancreatic cancer is associated with a strong desmoplastic reaction, detectable with Podoplanin (D2-40) [185]. While we observed a robust desmoplastic reaction in the submucosa, it was notably absent around the intramucosal cancer cells. The lamina propria also remained intact around these cells.

To quantify these differences, we employed a panel of 12 IHC markers. This panel distinguished between the two main pancreatic cancer subtypes, the classical and basal-like [186], of which the basal-like have a worse prognosis [187–189]. Additionally, we included markers for proliferation, general PDAC markers and tumor markers. These markers revealed significant reductions in basal-like PDAC markers within the intramucosal cells compared to the submucosal cells (Figure 15). Furthermore, the intestinal markers MUC2 and CK20 exhibited significant increases in the intramucosal cells. These changes indicated a remarkable shift in the PDAC cellular phenotype specific to their location.

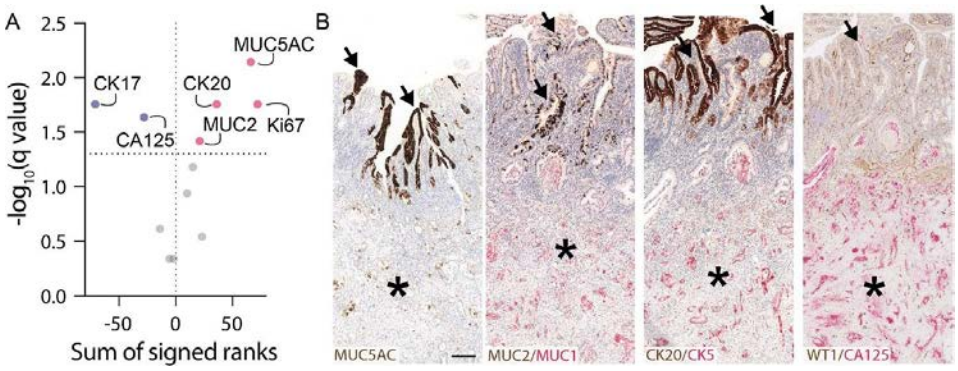

Figure 15. Shift in cancer cell phenotype on integration into the duodenal mucosa. (A) Volcano plot displaying the significant differences, red: higher in the mucosa; blue: higher in the submucosa. (B) Differences in IHC in mucosal vs. submucosal tumor cells for MUC5AC, MUC2/MUC1, CK20/CK5, and WT1/CA125. Arrows show cancer cells integrated into the mucosa, while asterisks mark the submucosa.

We also observed location-dependent changes in the proliferation of cancer cells. As the cancer cells infiltrated the intestinal crypts, we noted increased proliferation compared to the submucosal cancer cells. This increase paralleled the normally heightened proliferation of intestinal crypt cells, responsible for replacing cells in the villi. Additionally, we observed a gradual shift in the expression of normal intestinal markers CK20 and MUC2. CK20 is typically expressed in intestinal epithelial cells, while MUC2 is a marker for goblet cells. We demonstrated that as the cancer cells moved higher up the villi, they progressively increased their expression of these markers, even adopting a goblet cell phenotype (Figure 16).

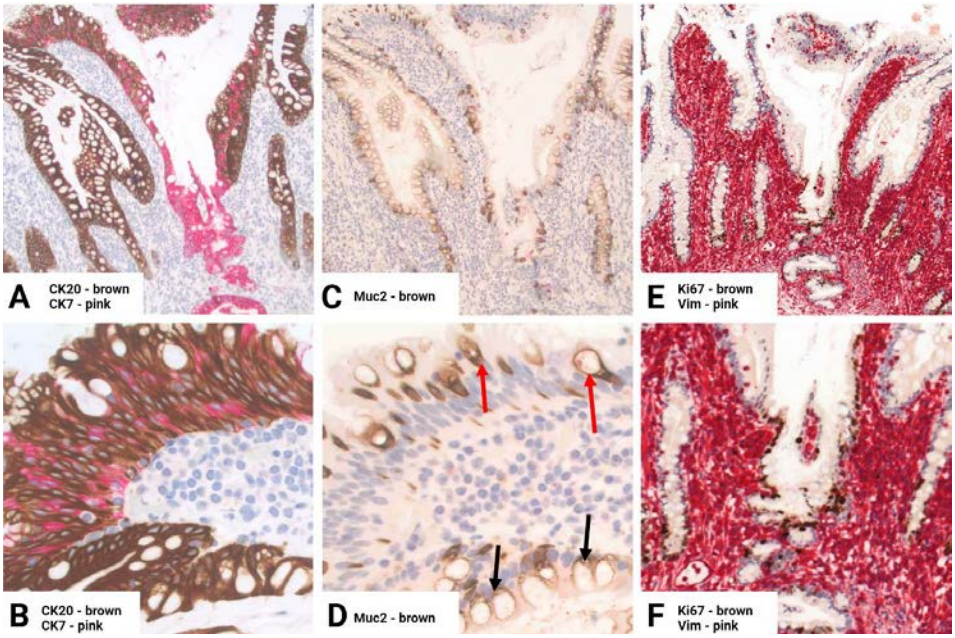

Figure 16. Location-dependent changes in the intestinal villi infiltrated by PDAC cells. All images are serial sections of the same tissue sample. The bottom row corresponds to a magnified image of a specific location from the top row. A and B) CK7 identifies PDAC cancer cells in the villus. A gradual change in CK20 positivity can be observed, similar to its expression in normal intestinal cells. C) Goblet cell differentiation visualized with Muc2 staining in cancer cells and normal cells. D) Black arrows indicate non-cancerous cells, while red arrows indicate cancer cells and show the difference in morphology between the two. E) Ki67 nuclear staining marks cells under proliferation, with a Vimentin staining of the stroma. F) Magnified image of the intestinal crypt where increased proliferation can be observed in cancer cells, compared to below or above the crypt.

In our animal model, the KPC mouse model, closely resembling human PDAC, driven by KRAS and Tp53 mutations [171], we observed a similar phenomenon as described in human tissues. In one of the six mice analyzed, PDAC cells infiltrated the intestinal mucosa. In this case, we analyzed HMGA2 expression, a marker for basal-like PDAC cells, and found it to be downregulated in the intra-mucosal cancer cells. These findings show the remarkable plasticity of cancer cells and to what extent they are influenced by the normal tissue environment, to the extent that even cellular proliferation can be changed depending on the localization of the cancer cells.

5 Conclusions

This thesis aimed to study the tumor microenvironment observed in human cancers. Two methods were developed, exploiting the information contained within the Human Protein Atlas (HPA) database, which provides data on protein expression in both normal and malignant tissues. Our research findings were validated through a combination of *in vivo* and *in vitro* experiments. Finally, by systematically analyzing histopathological data from patients with pancreatic cancer, we investigated to what extent the normal tissue exerts its influence on cancer cells.

The key findings from the studies included in this thesis are:

- **AtlasGrabber Software:** The AtlasGrabber software facilitates the systematic comparative analysis of the HPA database, enabling the identification of new IHC markers for various cell types and components within the ECM. It is especially useful for contrasting protein expression between normal and cancerous tissues (**paper V**).
- **New CAF markers:** 12 novel IHC markers of CAFs were identified, 4 of which were linked to Rho-kinase signaling pathways in cancer-associated fibroblasts. These provide new insights into cellular changes that occur in the cancer associated fibroblasts (**Paper I**).
- **Impact of RhoA in fibroblasts**: Fibroblasts lacking RhoA exhibited a diminished capacity to inhibit tumor growth *in vivo* and *in vitro*, highlighting the functional relevance of this protein in cancer invasion and metastasis (**Paper III**).
- **Decorin in the TME:** The proteoglycan Decorin of the ECM was found to be consistently downregulated in the tumor microenvironment and sheds light on the extensive matrix modulation and changes around cancers (**Paper II**).
- **Cell Plasticity in Pancreatic Ductal Carcinoma**: Pancreatic ductal carcinoma cells showed remarkable plasticity under the influence of the normal tissue environment — taking on a more indolent phenotype, their capacity to assume characteristics of normal intestinal cells, like goblet cell differentiation and affecting their proliferation. These findings highlight the remarkable influence that normal tissue can have on cancer cell phenotype (**Paper IV**).
- **Deep Learning Image Classifier:** Implementing a deep-learning-based image classifier, with image segmentation via color deconvolution and

binarization, enables the automatic, large-scale identification of analogous images in the HPA. This was shown by successfully identifying 44 previously undescribed markers for prostate basal cells (**Paper VI**).

The insights derived from these findings enhance our understanding of the morphological and molecular alterations occurring within the TME of cancers. The developed methodologies offer valuable tools for leveraging the HPA database to uncover further details about alterations within the TME, and our observations pertaining to the remarkable plasticity of cancer cells offer a novel perspective, potentially exploitable for new cancer treatment strategies.

6 Points of perspective

Expanding the HPA Analysis

Our initial analysis was confined to a limited subset of proteins derived from differential gene expression data between inhibitory and non-inhibitory fibroblasts. The validity of the gene set as a foundation for analysis remains unverified since it was not an aspect explored in our study. Utilizing the HPA, we could expand a similar investigation, performing analysis across the entire database for new signatures of CAFs and modifications in the ECM. Though this proposes a substantial and time-intensive endeavor, the method we developed in **paper VI** could make such a task possible, as the training image data from **paper I** is already available.

In **paper VI**, as a proof-of-concept, we examined prostate basal cells, chosen due to their distinct staining patterns, which form circular shapes around prostate glands. However, tumor tissues typically present more variation and complexity, mainly due to the variability in DAB staining in the images, sometimes staining the tumor and sometimes the stroma or both. This necessitates an initial segmentation into cancerous and non-cancerous tissues. This could be done by first performing a color deconvolution, then segmenting images to tumor and non-tumor regions based on the Hematoxylin staining, and then superimposing the images from the DAB staining, which should allow the analysis separately in tumor or non-tumor regions.

Using the HPA exclusively introduces an additional limitation as it encompasses only tissue microarrays, offering a restricted view of the tumor and tumor stroma. Analyzing the extent and uniformity of expression of identified proteins could be done by analyzing whole tissue slides.

Enhancements and Usability of the Deep Learning Classifier

Further exploration and enhancement, utilizing varied image processing techniques and newer deep learning models, are plausible future pursuits for our deep learning classifier. Alternatives such as excluding image binarization or incorporating deep learning for this step could be explored. Although we performed our analysis using Python code, user-friendly graphical interfaces, could expand its utility to researchers without coding expertise. Such apps could be designed to permit image uploads, enable CNN selection, and automatically

generate performance scores, thereby facilitating simpler and more efficient analyses.

The Cellular Plasticity of Cancers

The remarkable shift in cellular plasticity we observed in pancreatic cancers in **paper IV** raises the question of what factors in the normal tissues led to these changes. Presumably, these cues might mirror those directing stem cells in the intestinal crypts toward differentiation into various cells of the intestinal epithelium. Interestingly, the cues varied across different parts of the intestinal villi. Investigating potential differences in the normal microenvironment at different levels of the mucosal stroma and submucosa is feasible, with this information potentially being available in the HPA, and could, therefore, be explored with the methods we developed. In addition, the extent to which this plasticity is dependent on the cancer cells also remains unclear. Exploring tumors across a broader range that invade different tissues to ascertain if similar changes are observable could be valuable. However, this would necessitate clinical preparations, and tissue microarrays in the HPA database might prove insufficient. As in **paper III**, any potential findings from these observations in the HPA would need to be experimentally verified and tested to help us understand the underlying biology of these changes.

7 Acknowledgments

My thesis would not have been possible without the help and support of my family, friends, supervisors, and research collaborators. I want to extend my gratitude for their support throughout my doctoral studies.

To my main supervisor, **Ingemar Ernberg**, thank you for taking me on as a Ph.D. student, allowing me the freedom to pursue my research interests, and giving me constant support throughout my PhD. Your scientific curiosity, wide-spanning knowledge, and interests ranging from tumor biology to the humanities have given me a unique and memorable Ph.D. experience. I could not have wished for a better supervisor!

To **Georg Klein**, who first introduced me to science and awakened my interest in it. You gave me the confidence and opportunity to pursue a career in science. I will never forget your dedication to science and your support.

To **Laszlo Szekely**, who first gave the idea that formed the basis of my work and who first introduced me to image analysis. You are the sharpest scientist I know, and whenever I got stuck in my research, you were always there to give suggestions on how to proceed and with new ideas.

To my research colleagues and friends, **Twana**, with whom I will never forget our days in the Klein group, your good humor, the laughs we shared, and your positive outlook on life. To **Iurii,** with whom we had a fruitful collaboration and hopefully can continue that path. To **Andrii**, thank you for your help in designing the software and inspiring me to learn computer coding. To **Carlos,** for our scientific collaboration and your friendship. To **Peter Csermely** for introducing me to the world of network science!

To all my friends from the **Doctoral Student Association** and the **Summer School**, for having a memorable and fun PhD experience, especially **Susi**, **Maria**, and **Natasha,** for all the memorable parties we've been to and the many lunch dates! To my friend **Aafke** for being my bike partner throughout these years!

To my colleagues and friends in the Oncology Department, especially **Mark, Marcus, Una, Asaf, Georgios, Olof, Caroline,** and **Lisa**, for being good friends and making clinical work much more fun and rewarding. My residency supervisor, **Alex,** for helping me in my oncology training.

To my best friends, **Nedim**, **Aydin**, **JC**, **Peter, Shaun,** and **Gabi**. Each of you, in your unique way, has contributed to this journey, and I feel blessed to know that I always have you at my side.

To my father, **Bela,** who has inspired me with his work ethic devoted pursuit of knowledge and understanding of tumor biology, and for all the insights you have shared with me from your observations in histopathology.

To my mother, **Ildiko**, for your unconditional love and the sacrifices you have made throughout my life. For all the help and encouragement in my education and constantly pushing me to do better.

To my brothers **Lolo** and **Freddy** for always being kind, loving, and supportive. You are the best brothers anyone could have.

And mostly thanks to my sister **Felicia**, the closest and most important person in my life. For always knowing that you have my back. I could not have wished for a cooler and more loving sibling to grow up with!

8 References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA: A Cancer J Clin. 2022;72:7–33.

2. Hanahan D. Rethinking the war on cancer. The Lancet. 2014;383:558–63.

3. WHO. Cancer Fact Sheet. 2017. http://www.who.int/mediacentre/factsheets/fs297/en/.

4. Hanahan D, Coussens L. Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment. Cancer Cell. 2012;21:309–22.

5. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100:57–70.

6. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell. 2011;144:646–74.

7. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022;12:31–46.

8. Wolchok JD, Kluger H, Callahan MK, Postow MA, Rizvi NA, Lesokhin AM, et al. Nivolumab plus Ipilimumab in Advanced Melanoma. New England Journal of Medicine. 2013;369:122–33.

9. Rosenberg SA. Raising the Bar: The Curative Potential of Human Cancer Immunotherapy. Science Translational Medicine. 2012;4:127ps8-127ps8.

10. Lackner MR, Wilson TR, Settleman J. Mechanisms of acquired resistance to targeted cancer therapies. Future Oncology. 2012;8:999–1014.

11. Magee JA, Piskounova E, Morrison SJ. Cancer Stem Cells : Impact , Heterogeneity , and Uncertainty. Cancer Cell. 2012;21:283–96.

12. Fisher R, Pusztai L, Swanton C. Cancer heterogeneity: implications for targeted therapeutics. British Journal of Cancer. 2013;108:479–85.

13. DeVita VT, DeVita-Raeburn E. The death of cancer : after fifty years on the front lines of medicine, a pioneering oncologist reveals why the war on cancer is winnable--and how we can get there.

14. Flaherty KT, Infante JR, Daud A, Gonzalez R, Kefford RF, Sosman J, et al. Combined BRAF and MEK Inhibition in Melanoma with BRAF V600 Mutations. New England Journal of Medicine. 2012;367:1694–703.

15. Rehman FL, Lord CJ, Ashworth A. Synthetic lethal approaches to breast cancer therapy. Nature Reviews Clinical Oncology. 2010;7:718–24.

16. Chan DA, Giaccia AJ. Harnessing synthetic lethal interactions in anticancer drug discovery. Nature Reviews Drug Discovery. 2011;10:351–64.

17. Coussens LM, Werb Z. Inflammation and cancer. Nature. 2002;420:860–7.

18. Alison MR, Lin W-R, Lim SML, Nicholson LJ. Cancer stem cells: In the line of fire. Cancer Treatment Reviews. 2012;38:589–98.

19. Tomasek JJ, Gabbiani G, Hinz B, Chaponnier C, Brown RA. Myofibroblasts and mechano-regulation of connective tissue remodelling. Nature Reviews Molecular Cell Biology. 2002;3:349–63.

20. Kalluri R, Zeisberg M. Fibroblasts in cancer. Nature Reviews Cancer. 2006;6:392–401.

21. Olumi AF, Grossfeld GD, Hayward SW, Carroll PR, Tlsty TD, Cunha GR. Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium. Cancer research. 1999;59:5002–11.

22. Dumont N, Liu B, Defilippis RA, Chang H, Rabban JT, Karnezis AN, et al. Breast fibroblasts modulate early dissemination, tumorigenesis, and metastasis through alteration of extracellular matrix characteristics. Neoplasia (New York, NY). 2013;15:249–62.

23. Direkze NC, Hodivala-Dilke K, Jeffery R, Hunt T, Poulsom R, Oukrif D, et al. Bone Marrow Contribution to Tumor-Associated Myofibroblasts and Fibroblasts. Cancer Research. 2004;64:8492–5.

24. Zeisberg EM, Potenta S, Xie L, Zeisberg M, Kalluri R. Discovery of Endothelial to Mesenchymal Transition as a Source for Carcinoma-Associated Fibroblasts. Cancer Research. 2007;67:10123–8.

25. Petersen OW, Nielsen HL, Gudjonsson T, Villadsen R, Rank F, Niebuhr E, et al. Epithelial to Mesenchymal Transition in Human Breast Cancer Can Provide a Nonmalignant Stroma. The American Journal of Pathology. 2003;162:391–402.

26. Orr B, Riddick ACP, Stewart GD, Anderson RA, Franco OE, Hayward SW, et al. Identification of stromally expressed molecules in the prostate by tag-profiling of cancer-associated fibroblasts, normal fibroblasts and fetal prostate. Oncogene. 2012;31:1130–42.

27. Marsh T, Pietras K, McAllister SS. Fibroblasts as architects of cancer pathogenesis. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2013;1832:1070–8.

28. Fukumura D, Xavier R, Sugiura T, Chen Y, Park EC, Lu N, et al. Tumor induction of VEGF promoter activity in stromal cells. Cell. 1998;94:715–25.

29. Erez N, Truitt M, Olson P, Hanahan D, Hanahan D. Cancer-Associated Fibroblasts Are Activated in Incipient Neoplasia to Orchestrate Tumor-Promoting Inflammation in an NF-κB-Dependent Manner. Cancer Cell. 2010;17:135–47.

30. Ozbek S, Balasubramanian PG, Chiquet-Ehrismann R, Tucker RP, Adams JC. The Evolution of Extracellular Matrix. Molecular Biology of the Cell. 2010;21:4300–5.

31. Hynes RO, Naba A. Overview of the matrisome-An inventory of extracellular matrix constituents and functions. Csh Perspect Biol. 2012;4:a004903.

32. Myllyharju J, Kivirikko KI. Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends in Genetics. 2004;20:33–43.

33. Byron A, Humphries JD, Humphries MJ. Defining the extracellular matrix using proteomics. International Journal of Experimental Pathology. 2013;94:75–92.

34. Frantz C, Stewart KM, Weaver VM. The extracellular matrix at a glance. J Cell Sci. 2010;123:4195–200.

35. Sorokin L. The impact of the extracellular matrix on inflammation. Nature Reviews Immunology. 2010;10:712–23.

36. Levental KR, Yu H, Kass L, Lakins JN, Egeblad M, Erler JT, et al. Matrix Crosslinking Forces Tumor Progression by Enhancing Integrin Signaling. Cell. 2009;139:891–906.

37. Barker HE, Cox TR, Erler JT. The rationale for targeting the LOX family in cancer. Nature Reviews Cancer. 2012;12:540–52.

38. Lu P, Weaver VM, Werb Z. The extracellular matrix: A dynamic niche in cancer progression. The Journal of Cell Biology. 2012;196:395–406.

39. Pirog-Garcia KA, Meadows RS, Knowles L, Heinegard D, Thornton DJ, Kadler KE, et al. Reduced cell proliferation and increased apoptosis are significant pathological mechanisms in a murine model of mild pseudoachondroplasia resulting from a mutation in the C-terminal domain of COMP. Human Molecular Genetics. 2007;16:2072–88.

40. Fresquet M, Jackson GC, Loughlin J, Briggs MD. Novel mutations in exon 2 of *MATN3* affect residues within the α-helices of the A-domain and can result in the intracellular retention of mutant matrilin-3. Human Mutation. 2008;29:330–330.

41. Bergamaschi A, Tagliabue E, S?rlie T, Naume B, Triulzi T, Orlandi R, et al. Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome. The Journal of Pathology. 2008;214:357–67.

42. Ruoslahti E. Structure and Biology of Proteoglycans. Annu Rev Cell Dev Biol. 1988;4:229–55.

43. Krusius T, Ruoslahti E. Primary structure of an extracellular matrix proteoglycan core protein deduced from cloned cDNA. Proc Natl Acad Sci. 1986;83:7683–7.

44. Theocharis AD, Skandalis SS, Tzanakakis GN, Karamanos NK. Proteoglycans in health and disease: novel roles for proteoglycans in malignancy and their pharmacological targeting. FEBS J. 2010;277:3904–23.

45. Danielson KG, Baribault H, Holmes DF, Graham H, Kadler KE, Iozzo RV. Targeted Disruption of Decorin Leads to Abnormal Collagen Fibril Morphology and Skin Fragility. J Cell Biol. 1997;136:729–43.

46. Keene DR, Antonio JDS, Mayne R, McQuillan DJ, Sarris G, Santoro SA, et al. Decorin Binds Near the C Terminus of Type I Collagen*. J Biol Chem. 2000;275:21801–4.

47. Zhang G, Ezura Y, Chervoneva I, Robinson PS, Beason DP, Carine ET, et al. Decorin regulates assembly of collagen fibrils and acquisition of biomechanical properties during tendon development. J Cell Biochem. 2006;98:1436–49.

48. Zhang G, Chen S, Goldoni S, Calder BW, Simpson HC, Owens RT, et al. Genetic Evidence for the Coordinated Regulation of Collagen Fibrillogenesis in the Cornea by Decorin and Biglycan*. J Biol Chem. 2009;284:8888–97.

49. Merline R, Moreth K, Beckmann J, Nastase MV, Zeng-Brouwers J, Tralhão JG, et al. Signaling by the Matrix Proteoglycan Decorin Controls Inflammation and Cancer Through PDCD4 and MicroRNA-21. Sci Signal. 2011;4:ra75.

50. Iozzo RV, Moscatello DK, McQuillan DJ, Eichstetter I. Decorin is a biological ligand for the epidermal growth factor receptor. The Journal of biological chemistry. 1999;274:4489–92.

51. Santra M, Reed CC, Iozzo RV. Decorin binds to a narrow region of the epidermal growth factor (EGF) receptor, partially overlapping but distinct from the EGF-binding epitope. The Journal of biological chemistry. 2002;277:35671–81.

52. Iozzo RV, Buraschi S, Genua M, Xu S-Q, Solomides CC, Peiper SC, et al. Decorin antagonizes IGF receptor I (IGF-IR) function by interfering with IGF-IR

activity and attenuating downstream signaling. The Journal of biological chemistry. 2011;286:34712–21.

53. Goldoni S, Humphries A, Nyström A, Sattar S, Owens RT, McQuillan DJ, et al. Decorin is a novel antagonistic ligand of the Met receptor. The Journal of cell biology. 2009;185:743–54.

54. Neill T, Schaefer L, Iozzo RV. Decorin: a guardian from the matrix. The American journal of pathology. 2012;181:380–7.

55. Iozzo RV, Chakrani F, Perrotti D, McQuillan DJ, Skorski T, Calabretta B, et al. Cooperative action of germ-line mutations in decorin and p53 accelerates lymphoma tumorigenesis. Proceedings of the National Academy of Sciences of the United States of America. 1999;96:3092–7.

56. Bi X, Tong C, Dockendorff A, Bancroft L, Gallagher L, Guzman G, et al. Genetic deficiency of decorin causes intestinal tumor formation through disruption of intestinal cell maturation. Carcinogenesis. 2008;29:1435–40.

57. Coulson-Thomas VJ, Gesteira TF, Coulson-Thomas YM, Vicente CM, Tersariol ILS, Nader HB, et al. Fibroblast and prostate tumor cell cross-talk: fibroblast differentiation, TGF-β, and extracellular matrix down-regulation. Experimental cell research. 2010;316:3207–26.

58. McDoniels-Silvers AL, Nimri CF, Stoner GD, Lubet RA, You M. Differential gene expression in human lung adenocarcinomas and squamous cell carcinomas. Clinical cancer research : an official journal of the American Association for Cancer Research. 2002;8:1127–38.

59. Oda G, Sato T, Ishikawa T, Kawachi H, Nakagawa T, Kuwayama T, et al. Significance of stromal decorin expression during the progression of breast cancer. Oncology reports. 2012;28:2003–8.

60. Crnogorac-Jurcevic T, Efthimiou E, Capelli P, Blaveri E, Baron A, Terris B, et al. Gene expression profiles of pancreatic cancer and stromal desmoplasia. Oncogene. 2001;20:7437–46.

61. Köninger J, Giese T, Mola FF di, Wente MN, Esposito I, Bachem MG, et al. Pancreatic tumor cells influence the composition of the extracellular matrix. Biochemical and biophysical research communications. 2004;322:943–9.

62. Salomäki HH, Sainio AO, Söderström M, Pakkanen S, Laine J, Järveläinen HT. Differential expression of decorin by human malignant and benign vascular tumors. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society. 2008;56:639–46.

63. Csordás G, Santra M, Reed CC, Eichstetter I, McQuillan DJ, Gross D, et al. Sustained down-regulation of the epidermal growth factor receptor by decorin. A mechanism for controlling tumor growth in vivo. The Journal of biological chemistry. 2000;275:32879–87.

64. Baghy K, Iozzo RV, Kovalszky I. Decorin-TGFβ axis in hepatic fibrosis and cirrhosis. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society. 2012;60:262–8.

65. Kiefer J a, Farach-Carson MC. Type I collagen-mediated proliferation of PC3 prostate carcinoma cell line: implications for enhanced growth in the bone microenvironment. Matrix biology : journal of the International Society for Matrix Biology. 2001;20:429–37.

66. Armstrong T, Packham G, Murphy LB, Bateman AC, Conti JA, Fine DR, et al. Type I collagen promotes the malignant phenotype of pancreatic ductal adenocarcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research. 2004;10:7427–37.

67. Koenig A, Mueller C, Hasel C, Adler G, Menke A. Collagen type I induces disruption of E-cadherin-mediated cell-cell contacts and promotes proliferation of pancreatic carcinoma cells. Cancer research. 2006;66:4662–71.

68. Provenzano PP, Inman DR, Eliceiri KW, Knittel JG, Yan L, Rueden CT, et al. Collagen density promotes mammary tumor initiation and progression. BMC medicine. 2008;6:11.

69. Mani SA, Guo W, Liao M-J, Eaton ENg, Ayyanan A, Zhou AY, et al. The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. Cell. 2008;133:704–15.

70. Gao D, Joshi N, Choi H, Ryu S, Hahn M, Catena R, et al. Myeloid Progenitor Cells in the Premetastatic Lung Promote Metastases by Inducing Mesenchymal to Epithelial Transition. Cancer Research. 2012;72:1384–94.

71. Chao Y, Wu Q, Acquafondata M, Dhir R, Wells A. Partial Mesenchymal to Epithelial Reverting Transition in Breast and Prostate Cancer Metastases. Cancer Microenvironment. 2012;5:19–28.

72. Chaffer CL, Thompson EW, Williams ED. Mesenchymal to Epithelial Transition in Development and Disease. Cells Tissues Organs. 2007;185:7–19.

73. Thiery JP, Acloque H, Huang RYJ, Nieto MA. Epithelial-Mesenchymal Transitions in Development and Disease. Cell. 2009;139:871–90.

74. Lotem J, Sachs L. Different blocks in the differentiation of myeloid leukemic cells. Proceedings of the National Academy of Sciences of the United States of America. 1974;71:3507–11.

75. Huang ME, Ye YC, Chen SR, Chai JR, Lu JX, Zhoa L, et al. Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia. Blood. 1988;72:567–72.

76. Chomienne C, Ballerini P, Balitrand N, Daniel MT, Fenaux P, Castaigne S, et al. All-trans retinoic acid in acute promyelocytic leukemias. II. In vitro studies: structure-function relationship. Blood. 1990;76:1710–7.

77. Stoker MG, Shearer M, O'Neill C. Growth inhibition of polyoma-transformed cells by contact with static normal fibroblasts. Journal of cell science. 1966;1:297–310.

78. Partanen JI, Nieminen AI, Mäkelä TP, Klefstrom J. Suppression of oncogenic properties of c-Myc by LKB1-controlled epithelial organization. Proceedings of the National Academy of Sciences of the United States of America. 2007;104:14694–9.

79. Partanen JI, Nieminen AI, Klefstrom J. 3D view to tumor suppression: lkb1, polarity and the arrest of oncogenic c-myc. Cell Cycle. 2009;8:716–24.

80. Howlett AR, Petersen OW, Steeg PS, Bissell MJ. A novel function for the nm23-H1 gene: overexpression in human breast carcinoma cells leads to the formation of basement membrane and growth arrest. Journal of the National Cancer Institute. 1994;86:1838–44.

81. Ingber DE, Madri JA, Jamieson JD. Role of basal lamina in neoplastic disorganization of tissue architecture. Proceedings of the National Academy of Sciences of the United States of America. 1981;78:3901–5.

82. Ingber DE, Madri JA, Jamieson JD. Basement membrane as a spatial organizer of polarized epithelia. Exogenous basement membrane reorients pancreatic epithelial tumor cells in vitro. The American journal of pathology. 1986;122:129–39.

83. Watanabe TK, Hansen LJ, Reddy NK, Kanwar YS, Reddy JK. Differentiation of pancreatic acinar carcinoma cells cultured on rat testicular seminiferous tubular basement membranes. Cancer research. 1984;44:5361–8.

84. Weaver VM, Petersen OW, Wang F, Larabell CA, Briand P, Damsky C, et al. Reversion of the malignant phenotype of human breast cells in three-dimensional culture and in vivo by integrin blocking antibodies. The Journal of cell biology. 1997;137:231–45.

85. Bissell MJ, Hines WC. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. Nat Med. 2011;17:320–9.

86. Dolberg DS, Bissell MJ. Inability of Rous sarcoma virus to cause sarcomas in the avian embryo. Nature. 309:552–6.

87. Stoker AW, Hatier C, Bissell MJ. The embryonic environment strongly attenuates v-src oncogenesis in mesenchymal and epithelial tissues, but not in endothelia. The Journal of cell biology. 1990;111:217–28.

88. Zijl F van, Mair M, Csiszar A, Schneller D, Zulehner G, Huber H, et al. Hepatic tumor?stroma crosstalk guides epithelial to mesenchymal transition at the tumor edge. Oncogene. 2009;28:4022–33.

89. Condeelis J, Pollard JW. Macrophages: Obligate Partners for Tumor Cell Migration, Invasion, and Metastasis. Cell. 2006;124:263–6.

90. Gocheva V, Wang H-W, Gadea BB, Shree T, Hunter KE, Garfall AL, et al. IL-4 induces cathepsin protease activity in tumor-associated macrophages to promote cancer growth and invasion. Genes & Development. 2010;24:241–55.

91. Condeelis J, Segall JE. Intravital imaging of cell movement in tumours. Nature Reviews Cancer. 2003;3:921–30.

92. Wyckoff JB, Wang Y, Lin EY, Li J -f., Goswami S, Stanley ER, et al. Direct Visualization of Macrophage-Assisted Tumor Cell Intravasation in Mammary Tumors. Cancer Research. 2007;67:2649–56.

93. Carmeliet P, Jain RK. Molecular mechanisms and clinical applications of angiogenesis. Nature. 2011;473:298–307.

94. Coons AH, Creech HJ, Jones RN. Immunological Properties of an Antibody Containing a Fluorescent Group.*. Proc Soc Exp Biol Med. 1941;47:200–2.

95. Ramos-Vara JA. Technical Aspects of Immunohistochemistry. Vet Pathol. 2005;42:405–26.

96. Nakane PK, Pierce GB. Enzyme-labeled antibodies for the light and electron microscopic localization of tissue antigens. J Cell Biol. 1967;33:307–18.

97. Dybdal N, Leiberman G, Anderson S, McCune B, Bajamonde A, Cohen RL, et al. Determination of HER2 Gene Amplification by Fluorescence In situ Hybridization and Concordance with the Clinical Trials Immunohistochemical Assay in Women with Metastatic Breast Cancer Evaluated for Treatment with Trastuzumab. Breast Cancer Res Treat. 2005;93:3–11.

98. Muss HB, Thor AD, Berry DA, Kute T, Liu ET, Koerner F, et al. C-erbB-2 Expression and Response to Adjuvant Therapy in Women with Node-Positive Early Breast Cancer. N Engl J Med. 1994;330:1260–6.

99. Fitzgibbons PL, Page DL, Weaver D, Thor AD, Allred DC, Clark GM, et al. Prognostic Factors in Breast Cancer. Arch Pathol Lab Med. 2000;124:966–78.

100. Hashimoto K, Sasajima Y, Ando M, Yonemori K, Hirakawa A, Furuta K, et al. Immunohistochemical Profile for Unknown Primary Adenocarcinoma. PLoS ONE. 2012;7:e31181.

101. Selves J, Long-Mira E, Mathieu M-C, Rochaix P, Ilié M. Immunohistochemistry for Diagnosis of Metastatic Carcinomas of Unknown Primary Site. Cancers. 2018;10:108.

102. Taylor C. The nature of reed-sternberg cells and other malignant " reticulum " cells. Lancet. 1974;304:802–7.

103. Taylor CR, Burns J. The demonstration of plasma cells and other immunoglobulin-containing cells in formalin-fixed, paraffin-embedded tissues using peroxidase-labelled antibody. J Clin Pathol. 1974;27:14.

104. Taylor CR, Mason DY. The immunohistological detection of intracellular immunoglobulin in formalin-paraffin sections from multiple myeloma and related conditions using the immunoperoxidase technique. Clin Exp Immunol. 1974;18:417–29.

105. Shi SR, Key ME, Kalra KL. Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. J Histochem Cytochem : Off J Histochem Soc. 1991;39:741–8.

106. KÖHLER G, MILSTEIN C. Continuous cultures of fused cells secreting antibody of predefined specificity. Nature. 1975;256:495–7.

107. McMichael AJ, Pilch JR, Fabre JW, Mason DY, Galfré G, Milstein C. A human thymocyte antigen defined by a hybrid myeloma monoclonal antibody. Eur J Immunol. 1979;9:205–10.

108. Immunohistochemistry. https://www.proteinatlas.org/learn/method/immunohistochemistry. Accessed 6 Oct 2023.

109. Sternberger LA, Hardy PH, Cuculis JohnJ, Meyer HG. The unlabeled antibody enzyme method of immunohistochemistry preparation and properties of soluble antigen-antibody complex (horseradish peroxidase-

antihorseradish peroxidase) and its use in identification of spirochetes. J Histochem Cytochem. 1969;18:315–33.

110. Taylor CR, Rudbeck L. Immunohistochemical Staining Methods - IHC Guidebook. Agilent Dako.

111. Loos CM van der. Multiple immunoenzyme staining: methods and visualizations for the observation with spectral imaging. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society. 2008;56:313–28.

112. Cottrell TR, Taube JM. PD-L1 and Emerging Biomarkers in Immune Checkpoint Blockade Therapy. Cancer J. 2018;24:41–6.

113. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. N Engl J Med. 2016;375:1823–33.

114. Yi M, Jiao D, Xu H, Liu Q, Zhao W, Han X, et al. Biomarkers for predicting efficacy of PD-1/PD-L1 inhibitors. Mol Cancer. 2018;17:129.

115. Taylor CR, Levenson RM. Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment II. Histopathology. 2006;49:411–24.

116. Human Protein Atlas. 2015. http://www.proteinatlas.org/.

117. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science. 2015;347:1260419–1260419.

118. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based Human Protein Atlas. Nat Biotechnol. 2010;28:1248–50.

119. Pontén F, Schwenk JM, Asplund A, Edqvist P-HD. The Human Protein Atlas as a proteomic resource for biomarker discovery. Journal of internal medicine. 2011;270:428–46.

120. Antibody validation. 2017. http://www.proteinatlas.org/about/antibody+validation.

121. Uhlen M, Bandrowski A, Carr S, Edwards A, Ellenberg J, Lundberg E, et al. A proposal for validation of antibodies. Nature Methods. 2016;13:823–7.

122. UniProt. 2017. http://www.uniprot.org/.

123. Human Protein Atlas Primary Antibodies. https://www.atlasantibodies.com/primary-antibodies/triple-a-polyclonals/. Accessed 3 Oct 2023.

124. Billingsley FC. Processing Ranger and Mariner Photography. Opt Eng. 1966;4:404147-404147-.

125. Joblove GH, Greenberg D. Color spaces for computer graphics. ACM SIGGRAPH Comput Graph. 1978;12:20–5.

126. Ketcham DJ, Lowe RW, Weber JW. Image Enhancement Techniques for Cockpit Displays. 1974. https://doi.org/10.21236/ada014928.

127. Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, et al. Adaptive histogram equalization and its variations. Comput Vis, Graph, Image Process. 1987;39:355–68.

128. Otsu N. A Threshold Selection Method from Gray-Level Histograms. Ieee Transactions Syst Man Cybern. 1979;9:62–6.

129. Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. Anal Quantitative Cytol Histology Int Acad Cytol Am Soc Cytol. 2001;23:291–9.

130. Rueden CT, Schindelin J, Hiner MC, DeZonia BE, Walter AE, Arena ET, et al. ImageJ2: ImageJ for the next generation of scientific image data. Bmc Bioinformatics. 2017;18:529.

131. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: Open source software for digital pathology image analysis. Sci Rep. 2017;7:16878.

132. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. arXiv. 2012. https://doi.org/10.48550/arxiv.1201.0490.

133. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9:676–82.

134. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44.

135. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. 1943;5:115–33.

136. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychol Rev. 1958;65:386–408.

137. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323:533–6.

138. Linnainmaa S. Taylor expansion of the accumulated rounding error. BIT Numer Math. 1976;16:146–60.

139. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Comput. 1989;1:541–51.

140. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998;86:2278–324.

141. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017;60:84–90.

142. Bozoky B, Savchenko A. AtlasGrabber. https://github.com/b3nb0z/AtlasGrabber. Accessed 9 Dec 2020.

143. Foundation FS, Inc. GNU General Public License v. 3. 2007. https://www.gnu.org/licenses/gpl-3.0.en.html. Accessed 9 Dec 2020.

144. Data from the Human Protein Atlas in XML format. http://www.proteinatlas.org/about/download/proteinatlas.xml.gz. Accessed 13 Jun 2022.

145. Weinstein MH, Signoretti S, Loda M. Diagnostic utility of immunohistochemical staining for p63, a sensitive marker of prostatic basal cells. Modern Pathol. 2002;15:1302–8.

146. Brawer MK, Peehl DM, Stamey TA, Bostwick DG. Keratin Immunoreactivity in the Benign and Neoplastic Human Prostate. Cancer Res. 1985;45:3663–7.

147. Parsons JK, Gage WR, Nelson WG, Marzo AMD. p63 protein expression is rare in prostate adenocarcinoma: implications for cancer diagnosis and carcinogenesis. Urology. 2001;58:619–24.

148. Hedrick L, Epstein JI. Use of keratin 903 as an adjunct in the diagnosis of prostate carcinoma. Am J Surg Pathology. 1989;13:389–96.

149. Martin FJ, Amode MR, Aneja A, Austine-Orimoloye O, Azov AG, Barnes I, et al. Ensembl 2023. Nucleic acids Res. 2022;51:D933–41.

150. Flaberg E, Guven H, Savchenko A, Pavlova T, Kashuba V, Szekely L, et al. The architecture of fibroblast monolayers of different origin differentially influences tumor cell growth. International journal of cancer Journal international du cancer. 2012;131:2274–83.

151. Bodnar AG, Ouellette M, Frolkis M, Holt SE, Chiu C-P, Morin GB, et al. Extension of Life-Span by Introduction of Telomerase into Normal Human Cells. Science. 1998;279:349–52.

152. Consortium TU. The Universal Protein Resource (UniProt). Nucleic Acids Res. 2008;36 Database issue:D190–5.

153. Stark C, Breitkreutz B-J, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, et al. The BioGRID Interaction Database: 2011 update. Nucleic acids research. 2011;39:D698-704.

154. Prasad TSK, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, et al. Human Protein Reference Database--2009 update. Nucleic acids research. 2009;37:D767-72.

155. Bozoky B, Szekely L, Ernberg I, Savchenko A. AtlasGrabber: a software facilitating the high throughput analysis of the human protein atlas online database. Bmc Bioinformatics. 2022;23:546.

156. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science. 2013;339:819–23.

157. Kaighn ME, Narayan KS, Ohnuki Y, Lechner JF, Jones LW. Establishment and characterization of a human prostatic carcinoma cell line (PC-3). Investig Urol. 1979;17:16–23.

158. Flaberg E, Markasz L, Petranyi G, Stuber G, Dics?? F, Alchihabi N, et al. High-throughput live-cell imaging reveals differential inhibition of tumor cell proliferation by human fibroblasts. International Journal of Cancer. 2011;128:2793–802.

159. Sternberg undefined. Biomedical Image Processing. Computer. 1983;16:22–34.

160. Prewitt JMS, Mendelsohn ML. THE ANALYSIS OF CELL IMAGES*. Ann Ny Acad Sci. 2006;128:1035–53.

161. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. 2017.

162. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. 2016 IEEE Conf Comput Vis Pattern Recognit (CVPR). 2016;:2818–26.

163. Henry GH, Malewska A, Joseph DB, Malladi VS, Lee J, Torrealba J, et al. A Cellular Anatomy of the Normal Adult Human Prostate and Prostatic Urethra. Cell Rep. 2018;25:3530-3542.e5.

164. Schmitt T, Ogris C, Sonnhammer ELL. FunCoup 3.0: database of genome-wide functional coupling networks. Nucleic Acids Res. 2014;42:D380–8.

165. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019;47:D330–8.

166. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: Tool for the unification of biology. Nat Genet. 2000;25:25–9.

167. Schriml LM, Mitraka E, Munro J, Tauber B, Schor M, Nickle L, et al. Human Disease Ontology 2018 update: classification, content and workflow expansion. Nucleic Acids Res. 2019;47 Database issue:D955–62.

168. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur Ö, Anwar N, et al. Pathway Commons, a web resource for biological pathway data. Nucleic Acids Res. 2011;39 suppl_1:D685–90.

169. Rodchenkov I, Babur O, Luna A, Aksoy BA, Wong JV, Fong D, et al. Pathway Commons: 2019 Update. bioRxiv. 2019;:788521.

170. Moro CF, Fernandez-Woodbridge A, D'Souza MA, Zhang Q, Bozoky B, Kandaswamy SV, et al. Immunohistochemical typing of adenocarcinomas of the pancreatobiliary system improves diagnosis and prognostic stratification. PLoS ONE. 2016;11:e0166067.

171. Hingorani SR, Wang L, Multani AS, Combs C, Deramaudt TB, Hruban RH, et al. Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. Cancer Cell. 2005;7:469–83.

172. Strell C, Norberg KJ, Mezheyeuski A, Schnittert J, Kuninty PR, Moro CF, et al. Stroma-regulated HMGA2 is an independent prognostic marker in PDAC and AAC. Br J Cancer. 2017;117:65–77.

173. Cornish TC, Chakravarti A, Kapoor A, Halushka MK. HPASubC: A suite of tools for user subclassification of human protein atlas tissue images. J Pathol Inform. 2015;6:36.

174. Marr D, Hildreth E. Theory of edge detection. Proc R Soc Lond Ser B, Biol Sci. 1980;207:187–217.

175. Zhao X-H, Laschinger C, Arora P, Szászi K, Kapus A, McCulloch C a. Force activates smooth muscle alpha-actin promoter activity through the Rho signaling pathway. Journal of cell science. 2007;120:1801–9.

176. Moro CF, Bozóky B, Gerling M. Growth patterns of colorectal cancer liver metastases and their impact on prognosis: a systematic review. BMJ Open Gastroenterol. 2018;5:e000217.

177. Moro CF, Geyer N, Harrizi S, Hamidi Y, Söderqvist S, Kuznyecov D, et al. An idiosyncratic zonated stroma encapsulates desmoplastic liver metastases and originates from injured liver. Nat Commun. 2023;14:5024.

178. Latacz E, Höppener D, Bohlok A, Leduc S, Tabariès S, Moro CF, et al. Histopathological growth patterns of liver metastasis: updated consensus guidelines for pattern scoring, perspectives and recent mechanistic insights. Brit J Cancer. 2022;127:988–1013.

179. Payne SH. The utility of protein and mRNA correlation. Trends Biochem Sci. 2015;40:1–3.

180. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2012;13:227–32.

181. Edfors F, Danielsson F, Hallström BM, Käll L, Lundberg E, Pontén F, et al. Gene-specific correlation of RNA and protein levels in human cells and tissues. Mol Syst Biol. 2016;12:883.

182. Gry M, Rimini R, Strömberg S, Asplund A, Pontén F, Uhlén M, et al. Correlations between RNA and protein expression profiles in 23 human cell lines. BMC Genom. 2009;10:365.

183. Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, et al. Deep proteome and transcriptome mapping of a human cancer cell line. Mol Syst Biol. 2011;7:548–548.

184. Smith RA, Tang J, Tudur-Smith C, Neoptolemos JP, Ghaneh P. Meta-analysis of immunohistochemical prognostic markers in resected pancreatic cancer. Br J Cancer. 2011;104:1440–51.

185. Shindo K, Aishima S, Ohuchida K, Fujiwara K, Fujino M, Mizuuchi Y, et al. Podoplanin expression in cancer-associated fibroblasts enhances tumor progression of invasive ductal carcinoma of the pancreas. Mol Cancer. 2013;12:168–168.

186. Chan-Seng-Yue M, Kim JC, Wilson GW, Ng K, Figueroa EF, O'Kane GM, et al. Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. Nat Genet. 2020;52:231–40.

187. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SGH, Hoadley KA, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet. 2015;47:1168–78.

188. Maurer C, Holmstrom SR, He J, Laise P, Su T, Ahmed A, et al. Experimental microdissection enables functional harmonisation of pancreatic cancer subtypes. Gut. 2019;68:1034.

189. Puleo F, Nicolle R, Blum Y, Cros J, Marisa L, Demetter P, et al. Stratification of Pancreatic Ductal Adenocarcinomas Based on Tumor and Microenvironment Features. Gastroenterology. 2018;155:1999-2013.e3.