Proteomic Signatures of the Colorectal Cancer Desmoplastic Invasion Front

by

Georgios Karagiannis

Submitted in conformity with the requirements for the Degree of Doctor of Philosophy
Graduate Department of Laboratory Medicine and Pathobiology
University of Toronto

© Copyright by Georgios Karagiannis (2014)
Cancer-associated fibroblasts (CAFs) comprise a proponent compartment of the tumor microenvironment and are implicated in tumor progression. Local invasion and tumor-budding formation, i.e. detachment of individual cancer cells from the tumor invasion front, are regulated by CAFs and the desmoplastic stroma. Given accumulative evidence that such cells have undergone dedifferentiation and have acquired stem-like properties, it was hypothesized that they may be subjected to epithelial-to-mesenchymal transition (EMT) in a context-dependent manner. This thesis attempts to identify and characterize the contextual signals originating from the adjacent CAF population, responsible for mediating these malignant phenomena.

Colorectal cancer was opted as model system to address these questions. Co-cultures mimicking the colorectal cancer desmoplastic tumor-host cell interface were generated. Mass spectrometry-based proteomic methods and bioinformatics pipelines were deployed to draw comprehensive signatures of growth factors/mediators secreted in these cocultures, with potential of regulating EMT in neoplastic invasive margins. Given previously-established, tumor-suppressive roles of the BMP pathway, a group of bone morphogenetic protein (BMP) antagonists, namely GREM1, FST and HTRA3, demonstrated attractive candidacy as potential EMT mediators. Using in vitro assays, this thesis further provided experimental insights that antagonist-dependent disruption of
the BMP pathway induces enhanced motility and triggers mesenchymal phenotype switching in colon cancer cells.

Finally, this thesis provided translational impact of these results in colon cancer patients. BMP antagonists, especially HTRA3 and GREM1, were overexpressed in desmoplastic stromata with tumor budding presence, and their expression significantly correlated with increased tumor-budding score. Additionally, GREM1 was more prominently expressed in invasion fronts with nuclear versus membranous beta-catenin expression.

Overall, the data provided here support a working model, whereby contextual BMP antagonism may drive EMT, local invasion and de novo formation of tumor buds from the desmoplastic invasion front. Unraveling this axis thoroughly might be critical for clarifying pharmaceutical opportunities and proposing targeted strategies to manipulate metastasis.
ACKNOWLEDGEMENTS

The work included in this thesis has been a rewarding, inspiring and fruitful experience, which would have definitely been impossible without those who stood by my side throughout the entire journey. Following the paradigm of this thesis, I would therefore like to acknowledge the importance of my own personal macroenvironment that made this work flourish and blossom.

Foremost, I would like to express my sincere appreciation and gratitude to my mentor and supervisor, Dr. Eleftherios P. Diamandis, who constantly guided me through my experiments and patiently supported my ideas and wishful thinking, no matter the many obstacles ahead. His excellent advice for the duration of my studies has definitely boosted the essential critical thinking of a mature scientist and further deepened my passion for research. Dr. Eleftherios Diamandis has been acknowledged by the scientific community as an influential scientist that created his own school of scientists, and who are all now disseminated throughout the world. I am honestly honored and particularly happy that I belong to this esteemed and well-regarded group. I will never forget his successful metaphor during my first meeting with him in his office, in order to discuss the specifics of my project; he ensured me that I would be the driver and he would be the co-driver for the following five years. I have to admit this was quite a road trip!

Further gratitude for people that not only provided additional mentorship, but endless and invaluable interpersonal support throughout my journey are: Dr. Katerina Angelopoulou, Dr. Theofilos Poutahidis and Dr. Athanasios Dinopoulos; all three are faculty members in the School of Veterinary Medicine, Aristotle University of Thessaloniki, from which I had graduated before starting my PhD at the University of Toronto. They all deserve a separate paragraph to be
devoted to them, for teaching me high standards and the ethical aspects of scientific research, and showing me how to think “like a scientist”.

I am also very grateful for the productive interaction I had with the members of my PhD committee, Drs. Herman Yeger and Bharati Bapat, Professors of our Department. Their advice and support were substantial in the completion of this thesis. I would additionally like to mention Dr. Stephen Angers, member of my PhD transfer committee, for his time and constructive comments, midway to the completion of this thesis. Along the same lines, I would also like to thank Drs. Ivan Blasutig and George Yousef, members of my PhD examination committee, as well as Dr. Emmanuel Petricoin, who was nominated as the external appraiser for this thesis.

Very special thanks need to be given to a significant number of collaborators that assisted me with their advice, expertise and time in various aspects of my experimental work. Among all those, I would like to particularly mention Drs. Robert Riddell, Richard Kirsch, Anne Treacy, David Schaeffer and Constantina Petraki, whose contribution in the immunohistochemical experiments presented in this thesis was scrupulous and thoroughly-designed. I would also like to mention Drs. Joe Minta and Gary Bader for a magnificent collaboration in an independent project, which introduced me to novel areas of research and helped me acquire the skill of adapting my knowledge in a broad spectrum of biological questions.

I would like to also deeply thank all members of the lab, for not only maintaining an academic, but also an excellent and cozy atmosphere within these boundaries. All these people have contributed their best to the team, to help me balance a strict and professional working environment with some daily jokes and teasing. They made me feel as part of a large family.
My friends outside the lab, and roommates for most part of the last five years, Manos, Yiannis, Tassos and Igor gave me unreserved support. The time we had together in both good and bad moments (which are inevitable in the PhD journey!) will remain unforgettable memories.

Lastly, I would like to thank my family, especially my parents, Makis and Soula, and my younger brother, Nikos, for their endless love, and the fact that they always believed in me, even more than I did in myself. No matter I have spent almost six years away from my hometown, being an international student in a Canada, a country that is almost half a planet distance from Greece, I always felt they were sitting right next to me. I love them deeply and will never stop doing so. I am very grateful to them for giving me their never-ending support to fulfill this PhD thesis, a true academic Ithaka, down to its very end!

*Keep Ithaka always in your mind
Arriving there is what you are destined for
But do not hurry the journey at all.
Better if it lasts for years,
so you are old by the time you reach the island
wealthy with all you have gained on the way,
not expecting Ithaka to make you rich.*

*Ithaka gave you a marvelous journey.
Without her you would not have set out.
She has nothing left to give you now.*

*And if you find her poor, Ithaka won’t have fooled you.
Wise as you will have become, so full of experience,
you will have understood by then what these Ithakas mean*

- Constantine Kavafy
# TABLE of CONTENTS

ABSTRACT ................................................................................................................................. ii

ACKNOWLEDGEMENTS ............................................................................................................ iv

LIST of TABLES ......................................................................................................................... xi

LIST of FIGURES ....................................................................................................................... xii

LIST of ABBREVIATIONS ........................................................................................................... xv

CHAPTER 1: INTRODUCTION ..................................................................................................... 1

1.1. TUMOR-FIBROBLAST CELL INTERACTIONS ................................................................. 2
   1.1.1. The Tumor Microenvironment .................................................................................. 2
   1.1.2. Cancer-Associated Fibroblasts and Tumor Desmoplasia ......................................... 9
   1.1.3. Interdigital Migration: A CAF-Mediated Viewpoint of the Metastatic Cascade ....... 15
   1.1.4. Tumor Budding ......................................................................................................... 19

1.2. ONCOPROTEOGENS AND CANCER SECRETOMICS .................................................. 20
   1.2.1. General Notions ...................................................................................................... 20
   1.2.2. Protein Secretion Pathways .................................................................................... 20
   1.2.3. Defining the Cancer Secretome ............................................................................. 21
   1.2.4. Sources of Cancer Secretomes ............................................................................. 24
   1.2.5. In Silico Purification of the Cancer Secretome ....................................................... 25
   1.2.6. The Heterotypic Overview of the Cancer Secretome ........................................... 28

1.3. THE BONE MORPHOGENETIC PROTEIN (BMP) PATHWAY ......................................... 31
   1.3.1. The BMP Family of Proteins: Ligands and Their Receptors .................................... 31
   1.3.2. The BMP Signal Transduction Pathway ................................................................. 32
   1.3.3. Regulation of the BMP pathway through BMP Antagonism .................................. 35
   1.3.4. Sequestration of BMPs in the Extracellular Matrix .............................................. 37
   1.3.5. Implications of BMP Pathway in Development and Healthy Tissue Homeostasis .... 38
   1.3.6. Implications of BMP Pathway in Cancer Development and Progression ............ 39

1.4. COLORECTAL CANCER .................................................................................................... 41
   1.4.1. General Information ............................................................................................... 41
   1.4.2. Biomarkers in Colorectal Cancer .......................................................................... 41
   1.4.3. Genetic Pathways and Molecular Subtypes of Colorectal Cancer ......................... 43
   1.4.4. Colon Cancer Cell Lines: Can They Actually Recapitulate the Disease? ................. 45

1.5. STUDY AIMS, RATIONALE, AND HYPOTHESIS ......................................................... 47
   1.5.1. Rationale ................................................................................................................ 47
   1.5.2. Hypothesis .............................................................................................................. 48
   1.5.3. Objectives and Study Pipeline ............................................................................... 49
CHAPTER 6: TOPOGRAPHIC AND MECHANISTIC ASSOCIATION BETWEEN GREMLIN-1 EXPRESSION AND EPITHELIAL-TO-MESENCHYMAL TRANSITION .....123

6.1. INTRODUCTION.................................................................................................................. 124

6.2. EXPERIMENTAL PROCEDURES...................................................................................... 125
   6.2.1. Cell Lines and Culture Conditions.............................................................................. 125
   6.2.2. Gene-Expression Analysis of EMT-related Genes, using Real-Time PCR.............. 125
   6.2.3 Alkaline Phosphatase Assay ....................................................................................... 125
   6.2.4. Morphological and Computer-Assisted Assessment of EMT................................ 126
   6.2.5. Under-the-Agarose Cell Migration Assay ................................................................. 127
   6.2.6. Retrieval of Archived Immunohistochemical Cases and GREM1-EMT Correlation Studies in Selected Immunohistochemical Contexts......................................................... 127

6.3. RESULTS.......................................................................................................................... 131
   6.3.1. Contextual Correlation of Peritumoral GREM1 Expression with Selected Contexts Exhibiting an EMT-Pertaining Phenotype........................................................................... 131
   6.3.2. The BMP7-GREM1 Axis Promotes the EMT Phenotype in Colon Cancer Cells in vitro. 135
   6.3.3. The BMP7-GREM1 Axis Shifts the E/N-Cadherin Switch and Induces the Overexpression of the EMT-Promoting Transcription Factor and EMT marker Snail-1 .......... 137
   6.3.4. The BMP7-GREM1 Axis Promotes Colon Cancer Cell Migration in vitro ............. 139

6.4. DISCUSSION..................................................................................................................... 142
   6.4.1. Contextual Association between GREM1 and EMT: Immunohistochemical Evidence .... 142
   6.4.2. Functional Association between GREM1 and EMT/Migration: Preliminary Evidence .... 143
   6.4.3. Conclusive Remarks ................................................................................................. 145

CHAPTER 7: DISCUSSION AND FUTURE DIRECTIONS......................................................... 146

7.1. SUMMARY AND CONCLUSIONS..................................................................................... 147
   7.1.1. Brief Overview ........................................................................................................ 147
   7.1.2. Key Elements of the Experimental Pipeline............................................................ 147
   7.1.3. Working Model .................................................................................................... 149

7.2. FUTURE PERSPECTIVES............................................................................................... 152

CHAPTER 8: REFERENCES ..................................................................................................... 156

CHAPTER 9: APPENDIX ......................................................................................................... 194
LIST of TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.</td>
<td>Molecular portrait of 24 commercially available, commonly used CRC cell lines</td>
<td>46</td>
</tr>
<tr>
<td>5.1.</td>
<td>Classifications of contexts of the discovery cohort (N=90)</td>
<td>102</td>
</tr>
<tr>
<td>5.2.</td>
<td>Signaling interactions in the tumor microenvironment</td>
<td>106</td>
</tr>
<tr>
<td>A</td>
<td>The desmoplastic protein dataset (list of proteins in alphabetical order)</td>
<td>195</td>
</tr>
<tr>
<td>B</td>
<td>Patient demographic data (validation cohort; N=38), used for correction in the logistic regression model</td>
<td>199</td>
</tr>
</tbody>
</table>
## LIST of FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.</td>
<td>The cells of the tumor microenvironment</td>
<td>3</td>
</tr>
<tr>
<td>1.2.</td>
<td>Signaling interactions in the tumor microenvironment</td>
<td>8</td>
</tr>
<tr>
<td>1.3.</td>
<td>Normal and activated fibroblasts found at wound-healing and tumor sites</td>
<td>11</td>
</tr>
<tr>
<td>1.4.</td>
<td>The paracrine milieu of pleiotropic CAF functions in the tumor microenvironment</td>
<td>13</td>
</tr>
<tr>
<td>1.5.</td>
<td>The tumor invasion front</td>
<td>17</td>
</tr>
<tr>
<td>1.6.</td>
<td>&quot;Interdigital migration,&quot; a working model of CAF-directed metastasis</td>
<td>18</td>
</tr>
<tr>
<td>1.7.</td>
<td>Heterotypic overview of the cancer secretome</td>
<td>23</td>
</tr>
<tr>
<td>1.8.</td>
<td>Bioinformatics tools for prediction of protein secretion pathways</td>
<td>27</td>
</tr>
<tr>
<td>1.9.</td>
<td>The BMP signaling cascades</td>
<td>34</td>
</tr>
<tr>
<td>1.10.</td>
<td>Hypothesis and study design</td>
<td>50</td>
</tr>
<tr>
<td>2.1.</td>
<td>Proteomic pipeline for cell line secretome analysis</td>
<td>55</td>
</tr>
<tr>
<td>2.2.</td>
<td>Secretome dendrogram</td>
<td>57</td>
</tr>
<tr>
<td>2.3.</td>
<td>Development of <em>in-vitro</em> cocultures, mimicking the CRC desmoplastic microenvironment</td>
<td>60</td>
</tr>
<tr>
<td>2.4.</td>
<td>Secretome analysis</td>
<td>62</td>
</tr>
<tr>
<td>2.5.</td>
<td>Robustness and quality control of mass spectrometry-based protein identification</td>
<td>64</td>
</tr>
<tr>
<td>2.6.</td>
<td>Immunohistochemical determination of invasion front and association with collagen type XII</td>
<td>67</td>
</tr>
<tr>
<td>3.1.</td>
<td>Two-layered bioinformatics strategy for the organization of the desmoplastic protein dataset</td>
<td>77</td>
</tr>
<tr>
<td>3.2.</td>
<td>Enrichment map profiling of the desmoplastic protein dataset</td>
<td>79</td>
</tr>
</tbody>
</table>
3.3. Pathway analysis for identification of molecular hubs

4.1. *In vitro* overexpression of BMP inhibitors, as a result of tumor host-cell interactions

4.2. Verification studies in a CRC patient cohort, using immunohistochemistry

5.1. Study design and context selection

5.2. Context selection and characterization

5.3. BMP inhibitor expression scoring system

5.4. Coordinated epithelial and stromal expression of multiple BMPIs in colorectal cancer desmoplastic invasion fronts

5.5. Microenvironmental polarity of BMP inhibitors in CRC desmoplastic invasion front

5.6. Epithelial expression of BMP inhibitor HTRA3 is differentially regulated between tumor-budding and tumor-budding-free desmoplastic invasion fronts

5.7. Epithelial expression of BMP inhibitor HTRA3 holds pre-invasive information for tumor budding formation in colorectal cancer desmoplastic invasion fronts

5.8. Validation experiments in an independent patient cohort

6.1. Occludin (OCLN) and beta-catenin (B-CAT) immunohistochemistry

6.2. Association of GREM1 expression with EMT in intramuscular invasion fronts

6.3. Association of GREM1 expression with EMT in isolated tumor nests

6.4. Morphological assessment of EMT phenotype regulation by the BMP7/GREM1 axis *in vitro*

6.5. Regulation of EMT markers by the BMP7/GREM1 axis in HT29 colon cancer cells *in vitro*

6.6. GREM1 stimulates migration of SW480/SW620 cells *in vitro* in a BMP7-dependent manner

7.1. Working model on the potential implication of BMP inhibitors, as contextual signals for mediating local invasion and tumor budding formation
<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Enriched GO terms for the genes found in the three TGF-beta-superfamily-associated clusters</td>
<td>200</td>
</tr>
<tr>
<td>B</td>
<td>Categorization of proteins belonging to the TGF-beta superfamily into group I (Ligands/Receptors) and group II (Regulators/Inhibitors)</td>
<td>201</td>
</tr>
<tr>
<td>C</td>
<td>In-built negative controls</td>
<td>202</td>
</tr>
<tr>
<td>D</td>
<td>Pattern of BMP inhibitor expression across four quartiles of increasing BMP antagonism</td>
<td>203</td>
</tr>
<tr>
<td>E</td>
<td>Alkaline phosphatase activity, following BMP7 treatment in various colon cancer cells lines</td>
<td>204</td>
</tr>
<tr>
<td>F</td>
<td>Optimization of the under-the-agarose cell migration assay</td>
<td>205</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>Ang</td>
<td>Angiopoietin</td>
<td></td>
</tr>
<tr>
<td>ActR</td>
<td>Activin receptor</td>
<td></td>
</tr>
<tr>
<td>ALP</td>
<td>Alkaline phosphatase</td>
<td></td>
</tr>
<tr>
<td>ATF2</td>
<td>Activating transcription factor-2</td>
<td></td>
</tr>
<tr>
<td>BAMBI</td>
<td>BMP and activin membrane-bound protein</td>
<td></td>
</tr>
<tr>
<td>BAX</td>
<td>Apoptosis regulator</td>
<td></td>
</tr>
<tr>
<td>BCL-2</td>
<td>B-cell lymphoma-2</td>
<td></td>
</tr>
<tr>
<td>BIS(1)</td>
<td>BMP-induced signaling complex</td>
<td></td>
</tr>
<tr>
<td>BMP(I)/R</td>
<td>Bone morphogenetic protein (inhibitor)/(receptor)</td>
<td></td>
</tr>
<tr>
<td>RAF</td>
<td>Rapidly accelerated fibrosarcoma</td>
<td></td>
</tr>
<tr>
<td>CA 19-9</td>
<td>Carbohydrate antigen 19-9</td>
<td></td>
</tr>
<tr>
<td>CAF</td>
<td>Cancer-associated fibroblast</td>
<td></td>
</tr>
<tr>
<td>h-CALD</td>
<td>H-caldesmon</td>
<td></td>
</tr>
<tr>
<td>beta-CAT</td>
<td>Beta-catenin</td>
<td></td>
</tr>
<tr>
<td>CAV1</td>
<td>Caveolin-1</td>
<td></td>
</tr>
<tr>
<td>CEA</td>
<td>Carcinoembryonic antigen</td>
<td></td>
</tr>
<tr>
<td>CIF</td>
<td>Cancer invasion front</td>
<td></td>
</tr>
<tr>
<td>CIMP</td>
<td>CpG island methylator phenotype</td>
<td></td>
</tr>
<tr>
<td>CIN</td>
<td>Chromosomal instability</td>
<td></td>
</tr>
<tr>
<td>CM</td>
<td>Conditioned media</td>
<td></td>
</tr>
<tr>
<td>CD31</td>
<td>Cluster of differentiation 31</td>
<td></td>
</tr>
<tr>
<td>CDCHO</td>
<td>Chemically-defined Chinese hamster ovary</td>
<td></td>
</tr>
<tr>
<td>CLDN11</td>
<td>Claudin-11</td>
<td></td>
</tr>
<tr>
<td>CNS</td>
<td>Central nervous system</td>
<td></td>
</tr>
<tr>
<td>COL12A1</td>
<td>Collagen type XII</td>
<td></td>
</tr>
<tr>
<td>CRC</td>
<td>Colorectal cancer</td>
<td></td>
</tr>
<tr>
<td>CSC</td>
<td>Cancer stem cell</td>
<td></td>
</tr>
<tr>
<td>CTL</td>
<td>Cytotoxic T-lymphocyte</td>
<td></td>
</tr>
<tr>
<td>DAB</td>
<td>Diaminobenzidine</td>
<td></td>
</tr>
<tr>
<td>DAN</td>
<td>Differential screening-selected gene aberrative in neuroblastoma</td>
<td></td>
</tr>
<tr>
<td>DDT</td>
<td>Ditheiotheitol</td>
<td></td>
</tr>
<tr>
<td>DFS</td>
<td>Disease-free survival</td>
<td></td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s modified eagle’s medium</td>
<td></td>
</tr>
<tr>
<td>DPD</td>
<td>Desmoplastic protein dataset</td>
<td></td>
</tr>
<tr>
<td>ECM</td>
<td>Extracellular matrix</td>
<td></td>
</tr>
<tr>
<td>EGF</td>
<td>Epidermal growth factor</td>
<td></td>
</tr>
<tr>
<td>EMT</td>
<td>Epithelial-to-mesenchymal transition</td>
<td></td>
</tr>
<tr>
<td>EndMT</td>
<td>Endothelial-to-mesenchymal transition</td>
<td></td>
</tr>
<tr>
<td>EPH</td>
<td>Ephrin</td>
<td></td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
<td></td>
</tr>
<tr>
<td>FACT</td>
<td>Fibril-associated collagen with interrupted triple helices</td>
<td></td>
</tr>
<tr>
<td>FAP</td>
<td>Familial adenomatous polyposis</td>
<td></td>
</tr>
<tr>
<td>FAP</td>
<td>Fibroblast activation protein</td>
<td></td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal bovine serum</td>
<td></td>
</tr>
</tbody>
</table>
FDR  false discovery rate
FGF1/2 fibroblast growth factor 1/2
FIT  fecal immunochemical test
FOBT  fecal occult blood test
FST(L) follistatin (-like)
GDF  growth differentiation factor
GI  gastrointestinal
GO  gene ontology
GPX5 glutathione peroxidase-5
GREM1 gremlin-1
HGF  hepatocyte growth factor
HGTB  high grade tumor budding
HL  Hosmer-Lemeshow
HNPCC  hereditary non-polyposis colon cancer
HTRA1/3 high-temperature requirement-alpha-1/3 (a.k.a. probable serine peptidase-1/3)
IBD  inflammatory bowel disease
IGF1 insulin-like growth factor-1
IL-1beta interleukin-1 beta
INHBA inhibin-alpha
JNK  c-Jun N-terminal kinase
JP  juvenile polyposis
KLK  kallikrein-related peptidase
KRAS  Kirsten rat sarcoma viral oncogene homolog
LAMA3 laminin-alpha-3
LAMB1 laminin-beta-1
LGTB  low grade tumor budding
LIMK1 LIM-domain kinase-1
LOX  lysyl-6-oxidase
LTBP1 latent transforming growth factor-binding protein-1
MAPK  mitogen-activated protein kinase
MCP1 monocyte chemotactic protein-1
MDSC myeloid-derived suppressor cell
MEM  minimum essential medium
MET  mesenchymal-to-epithelial transition
MGP  matrix gla protein
MMP2/9 matrix metalloproteinase-2/9
MMP3  stromelysin-1
MSI  microsatellite instability
NCI  national cancer institute
NF  normal fibroblast
NK  natural killer
NOG  noggin
p-NPP  para-nitrophenylphosphate
NSCLC non-small lung cell carcinoma
OCLN  occludin
OD  optical density
OPN  osteopondin
OS  overall survival
uPA  urokinase-type plasminogen activator
PBS  phosphate buffered saline
PDGF  platelet-derived growth factor
PECAM  platelet/endothelial cell adhesion molecule
PI3K  phosphoinositide 3-kinase
PIK3CA  phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha
PRDC  protein related to dan and cerberus
PTEN  phosphatase and tensin homolog
RER  replication error
ROC  receiver operating characteristic
ROR2  tyrosine-protein kinase transmembrane receptor
SDF1  stromal-derived factor-1
mSEPT9  methylated septin-9
alpha-SMA  alpha-smooth muscle actin
SMAD  (small) mothers against decapentaplegic
SOST  sclerostin
TAM  tumor-associated macrophage
TGFB(R)  transforming growth factor-beta (receptor)
TIL  tumor-infiltrating lymphocyte
TNC  tenascin-C
TNF-alpha  tumor necrosis factor-alpha
TSG  twisted gastrulation
UC  ulcerative colitis
USAG1  uterine sensitization associated gene
VCAN  versican
VEGF(R)  vascular endothelial growth factor (receptor)
VG1  vegetally localized protein-1
XIAP  X-linked inhibitor of apoptosis protein
“Because today we live in a society in which spurious realities are manufactured by the media, by governments, by big corporations, by religious groups, political groups... So I ask, in my writing, What is real? Because unceasingly we are bombarded with pseudo-realities manufactured by very sophisticated people using very sophisticated electronic mechanisms. I do not distrust their motives; I distrust their power. They have a lot of it. And it is an astonishing power: that of creating whole universes, universes of the mind. I ought to know. I do the same thing.”

— Philip K. Dick

“Good morning! And in case I don't see you, good afternoon, good evening, and good night!”

— Truman Burbank, the Truman Show
CHAPTER 1

Introduction

Sections of this chapter have been published in Molecular Oncology, Molecular Cancer Research, and Journal of Proteomics:


Copyright permission has been granted.
1.1. TUMOR-FIBROBLAST CELL INTERACTIONS

1.1.1. The Tumor Microenvironment

1.1.1.1. The hallmarks of cancer and the tumor microenvironment

In an important review article by Hanahan and Weinberg entitled “Hallmarks of Cancer”\(^1\), the authors had proposed six distinct molecular features [i.e. (i) Sustaining proliferative signaling, (ii) Evading growth suppressors, (iii) Resisting cell death, (iv) Enabling replicative immortality, (v) Inducing angiogenesis, and (vi) Activating invasion and metastasis], which together constituted an organized and logical framework for understanding the remarkable diversity of neoplastic diseases. Conceptual progress made during the past decade on mechanistic underpinnings and the essence of tumorigenesis, allowed the reformulation of this concept\(^2\), by adding two emerging hallmarks [i.e. (vii) Evading immune destruction, and (viii) Reprogramming energy metabolism], as well as certain enabling characteristics –such as tumor-promoting inflammation and genomic instability–, which are collectively rendered responsible for the acquisition of these hallmarks. The consensus driven from the description of these hallmarks is the notion that as normal cells evolve progressively to a neoplastic state, they acquire a succession of capabilities, and that the multistep process of this pathogenesis could be rationalized by the need of incipient cancer cells to acquire the traits that enable them to become tumorigenic and ultimately malignant\(^2\).

However, it is quite clear that tumors are more than insular masses of proliferating cancer cells. Instead, they are complex tissues composed of multiple distinct cell types that participate in heterotypic interactions with one another. The recruited normal cells, collectively referred to as “tumor-associated stroma”, are seen as active participants of tumorigenesis rather than passive bystanders\(^2,3\). Therefore, these stromal cells contribute to the development and expression of certain of the aforementioned hallmark capabilities. As such, the biology of the tumor is now understood from the perspective of the individual specialized cells within it, as well as the “tumor microenvironment” that they construct along the course of tumorigenesis (Fig. 1.1.)\(^2\). This depiction has entirely substituted the earlier “reductionist” viewpoint, which described tumors as a collection of homogenous cancer cells, whose entire biology could be conceived by elucidating cell-autonomous properties\(^1,2\).
FIGURE 1.1. The cells of the tumor microenvironment. Above: an illustration of all distinct cell types constituting most solid tumors. Both the parenchyma and stroma of tumors contain distinct cell types and subtypes that collectively enable tumor growth and progression. Below: an illustration of various microenvironments of tumors depending on context. The multiple stromal cell types create a succession of tumor microenvironments that change as tumors invade normal tissue and thereafter seed and colonize distant tissues. The abundance, histologic organization, and phenotypic characteristics of the stromal cell types, as well as of the extracellular matrix (hatched background), evolve during progression, thereby enabling primary, invasive, and then metastatic growth. The surrounding normal cells of the primary and metastatic sites, shown only schematically, likely also affect the character of the various neoplastic microenvironments [Illustration adapted from: Hanahan and Weinberg (2011)].
1.1.1.2. Major constituents of the tumor microenvironment

Observations on the tumor microenvironment arise from the study of carcinomas, in which the tumor parenchyma consists of neoplastic epithelial cells, a compartment entirely distinct from the mesenchymal cells of the tumor-associated stroma. Epithelial cancer cells are the foundation of the disease; they are capable of initiating tumors and driving tumor progression forward, carrying along all the oncogenic hallmarks that define cancer as a genetic disease. On the other side, the mesenchymal compartment of tumors (i.e., endothelial cells; pericytes; immune inflammatory cells; cancer-associated fibroblasts; stem and progenitor cells of the tumor stroma) is primarily responsible for constructing the context, through which the hallmark capabilities are successively acquired. For purposes of gaining a global perspective on the structure of tumor microenvironment, here we will focus on major advancements on the biology and molecular implication of each cell type in the course of disease.

Prominent among the stromal constituents are the cells forming the tumor-associated vasculature. In general, cancer-associated angiogenesis represents a prominent hallmark of all cancers and is regulated by a complex dynamic network of both promoting and suppressive factors, whose balance is presumed to shift the angiogenic switch towards hyper- or hypovascularized states, depending on the context. Prominent cells in these interaction are normal endothelial cells, positioned in the underlying, unaffected, parental vessels. Over the last decade, a vast plethora of interconnected signaling circuitries/pathways involving ligands of signal-transducing receptors displayed in the surface of endothelial cells, such as Notch, Neuropillin, Robo, C-X-C-motif receptors (CXCRs), and Ephrin-A/b (Eph-A/B), have been unraveled and added next to the more traditional angiogenesis regulators, namely the vascular endothelial growth factor (VEGF), Angiopoietin (Ang), as well as fibroblast growth factor (FGF) signaling pathways. This knowledge supports a reciprocal paracrine mechanism, whereby cancer cell- and/or stromal cell-derived pro- and antiangiogenic factors are constantly affecting the angiogenic switch, which ultimately promotes endothelial cell proliferation and assembly into tubes (vasculogenesis), as well as sprouting of new vessels from existing ones (angiogenesis). Recently, high-throughput studies demonstrated distinctive gene and protein expression signatures in tumor-associated endothelial cells and identified cell-surface markers displayed on the luminal surfaces of normal versus tumor-endothelial cells. Consequently, the recruited endothelium (also described as “cancer-associated endothelium”) is a biologically distinct tissue than the one observed during developmental angiogenesis under physiological conditions, and could be targeted for therapeutic regimes.

A prominent subset of endothelial cells in the tumor microenvironment, whose role in cancer progression is poorly understood, is the one forming the lymphatic vessels. A common pattern observed in solid
carcinomas is the collapsing and absence of functionality of intratumoral lymphatics, mainly due to high interstitial pressure. However, functional lymphatics occur in the tumor margins and, in general, the periphery of cancer cohorts invading the healthy stroma. The latter are also believed to serve as channels for the seeding of metastasis in the draining lymph nodes. Endothelial cells forming the peritumoral lymphatics may respond and signal through different ligand/receptor systems than those often found in the cancer-associated endothelium (i.e. VEGF-C instead of VEGF-A). Thus, the endothelial cells lining the peritumoral lymphatics are seen as a distinct compartment of the tumor microenvironment.

Other mesenchymal cells of the tumor microenvironment, also associated with cancer-associated endothelium, are the pericytes that wrap around the endothelial tubing of blood vessels with finger-like projections. In the normal tissue, pericytes provide paracrine support signals, such as the antiproliferative and stabilizing Ang-1, which binds to Tie2 receptors of the adjacent endothelium to promote quiescence. Pharmacological studies on the disruption of the molecular axis of pericyte recruitment (i.e. platelet-derived growth factor (PDGF) signaling) from pericyte progenitors have demonstrated significantly reduced pericyte coverage of the cancer-associated endothelium and parallel destabilization of the vascular integrity and function of tumor vessels. Pericytes also collaborate with endothelial cells to synthesize basement membrane that anchors both cell types, helping them withstand the hydrostatic pressure of blood flow. As such, tumors, whose endothelium is characterized by low pericyte coverage, are more prone to permit cancer cell intravasation, an intermediate step of the metastatic cascade that enables the hematogenous dissemination of cancer cells in distant tissues.

Evidence had early begun to accumulate that infiltration of the tumor microenvironment by immune cells promotes tumor progression. This could be traced back to the observation that tumors are portrayed as “wounds than never heal.” In the case of normal wound healing and infectious diseases, immune inflammatory cells infiltrate these microenvironments in a transient fashion, and they subsequently disappear, in contrast to their persistence in sites of chronic inflammation, where their presence is linked to severe pathologies, such as fibrosis, aberrant angiogenesis and neoplasia. Despite the broad diversity of distinct immune cell types, there is now clear consensus that tumor-promoting effects are elicited by certain macrophage subtypes, mast cells, neutrophils, and T- and B-lymphocytes. Collective knowledge places these tumor-infiltrating immune cells at the tumor invasive margins, whereby they elicit a paracrine signaling repertoire to sustain tumor-associated angiogenesis, cancer cell proliferation, invasion and metastasis. A thorough enumeration of molecular strategies deployed to achieve the aforementioned cancer hallmarks is beyond the scope of this thesis. Suffice is to say that prominent molecular players, including epidermal growth factor (EGF), VEGF, fibroblast growth factor-2 (FGF2), chemokines and cytokines that further amplify inflammatory responses, matrix
degrading enzymes like matrix-metalloproteinase-9 (MMP9) and heparenase, are among the primary mediators secreted by immune inflammatory cells \(^{44, 48, 51, 52}\).

Fibroblasts may comprise a variable proportion across the spectrum of most carcinomas, constituting in many cases the dominant cell population of the tumor stroma \(^{5, 53-60}\). In an exaggerated paradigm, the fibroblastic population in pancreatic cancers may comprise more than 90% of the overall tumor mass \(^{61, 62}\). However, fibroblasts recruited from neoplastic signals and participating in the tumor microenvironment resemble myofibroblasts \(^{53, 63}\). The latter are normal cellular elements of many mucosal surfaces and basic structural components of the periglandular sheaths. They are also basic components of granulation tissue with important roles in wound healing and chronic inflammation \(^{2, 64}\). Also known as “cancer-associated fibroblasts” (CAFs), these recruited myofibroblasts tend to aggregate and encircle carcinoma cells invading the adjacent normal tissues \(^{53}\). Tumor transplantation studies show that CAFs enhance cancer cell proliferation, angiogenesis, invasion, and metastasis; indeed, tumors formed in mice after transplanting cancer cells admixed with CAFs are more malignant than those formed by transplanting cancer cells alone or cancer cells with normal fibroblasts \(^{54, 56}\).

Finally, despite the fact that most stromal cell types of the tumor microenvironment are derived from adjacent quiescent populations—a quite obvious reservoir for recruitment, it has been extensively shown recently that the bone marrow may also serve as a key source of tumor-associated stromal cells \(^{65-69}\). Of note, bone marrow-derived mesenchymal stem cells may be present in the tumor microenvironment, either as undifferentiated stromal cells to provide important paracrine framework that differentiated cells cannot express, or even serving as CAF, endothelial or pericyte progenitors \(^{65-67}\). Interestingly, subsets of partially differentiated myeloid precursors -intermediaries between circulating cells of bone marrow origin and typically differentiated immune cells of normal/inflamed tissues- are also encountered in the tumor microenvironment \(^{44}\). For instance, a class of tumor-infiltrating myeloid cells called myeloid-derived suppressor cells (MDSCs), co-expressing the macrophage marker CD11b and the neutrophil marker Gr1, has been shown to suppress the activity of cytotoxic T-lymphocyte (CTL) and natural killer (NK) cells, thus allowing cancer cells to evade immune destruction \(^{70}\).

In conclusion, various lines of evidence demonstrate that the tumor-associated stromal cells supply and support the developing tumors by recruitment and proliferation of pre-existing stromal, progenitor and stem cells, either these are derived from the adjacent, quiescent, healthy tissue, or the bone-marrow.
1.1.3. Heterotypic signaling circuitries dictate the successive acquisition of cancer hallmarks

A simplistic depiction of the complicated paracrine circuitry, governing the intercellular aspect of tumor biology is charted as signaling interactions between neoplastic and stromal cells within the tumor and its dynamic extracellular matrix (ECM) that they collectively erect and remodel (Fig. 1.2.)\(^1, 2, 5, 42, 71, 72\). A complete depiction of the network of microenvironmental signaling interactions is certainly beyond our reach, as the great majority of signaling molecules and pathways remains underexplored. However, the signaling map presented here (Fig. 1.2.), as adapted from the “Hallmarks of Cancer”\(^2\), accurately demonstrates the organizing principle of how the various components of the tumor microenvironment communicate through reciprocal paracrine interactions (arrows) that constantly and dynamically occur among them. A few pivotal signaling factors and molecules that belong to well-characterized paracrine axes are shown (in green squares)\(^2\). These growth factors mostly represent “recruiting” factors/cytokines, produced by one cell type to elicit a molecular response (or set of responses) to another cell type within the microenvironment\(^2\). Importantly, this simple schematic (Fig. 1.2.) implies another dimension of complexity: both neoplastic cells and stromal cells may progressively change during the multistep transformation of normal tissues into high-grade malignancies, which may further reflect underlying changes in the signaling circuitries depicted in these cells, followed by the accumulation and hardwiring of the acquired cancer hallmarks into their genomes and/or proteomes\(^2, 5\).
FIGURE 1.2. Signaling interactions in the tumor microenvironment. The illustration demonstrates a simplistic representation of cell types (cancer cells and cancer-associated stromal cells) constituting the tumor microenvironment, and the reciprocal paracrine interactions (arrows) that constantly and dynamically occur among them. A few signaling factors and molecules that belong to well-characterized paracrine axes are shown, as examples. These growth factors represent “recruiting” factors, produced by one cell type to elicit a particular response to another cell type within the tumor microenvironment. Please refer to the original list of abbreviations for these molecules. [Illustration adapted from: Hanahan and Weinberg (2011)].
1.1.2. Cancer-Associated Fibroblasts and Tumor Desmoplasia

1.1.2.1. General notions

Fibroblasts are the non-vascular, non-epithelial and non-inflammatory cellular population, embedded within the fibrillar matrix of the connective tissue. They primarily deposit and remodel various ECM components, and participate in the regulation of epithelial differentiation, inflammation, and wound healing. Notable ECM components synthesized by fibroblasts are collagen types I, III and V, and fibronectin. They also contribute to the formation of basement membranes that support epithelial anchoring and maintenance, through secretion of collagen type IV and laminin. Further homeostatic support of adjacent epithelia is achieved through the paracrine and/or juxtacrine secretion and/or position, respectively, of growth and trophic factors. Finally, besides synthesizing ECM, fibroblasts are also responsible for remodeling it by secreting specialized proteolytic enzymes, such as MMPs.

During wound healing, fibroblasts obtain a diverse phenotype, which assists them in invading proximal lesions and generating ECM components to serve as a scaffold for the aggregation of multiple other cell types. As such, fibroblasts are considered as the main mediators of scar formation and tissue fibrosis. Fibroblasts that are isolated from the site of a healing wound or from fibrotic tissue are able to secrete higher levels of normal ECM constituents and may also proliferate faster than their normal counterparts isolated from healthy organs. This particular phenotype is referred to as ‘activation’. Once the wound is repaired, such activated fibroblasts are either restored to their initial quiescent phenotype or undergo a specialized cell death-like program, also known as nemosis. Activated fibroblasts may also be found at various tumor sites, where they remain perpetually activated. Although myofibroblasts are beneficial in wound healing, their persistence in the tumor microenvironment contributes to pathological fibrosis, known as desmoplasia or desmoplastic reaction. The understanding of the role of desmoplastic reaction in cancer progression is still evolving, but is now believed that certain subpopulations of cancer-associated fibroblasts (CAFs) are important promoters of tumor growth and progression.

1.1.2.2. The fibroblast-to-myofibroblast transdifferentiation program

It has been shown that normal fibroblasts (NFs) undergo a certain series of molecular events, which may be summarized as a transdifferentiation program that transforms them into myofibroblast-resembling cells, namely CAFs. As a general notion, the signals that mediate such transition are not fully clarified regardless that certain advances are made in the field (Fig. 1.3.) For instance, phenotypic features of
CAFs can be induced by TGF-beta *in vitro*, a cytokine that is known to mediate fibroblast activation in wound healing and organ fibrosis. Therefore, similar pathways could be responsible for the generation of the CAF stromal population in solid carcinomas. In most carcinomas with the exaggerated paradigm of pancreatic cancer, the tumor stroma comprises the bigger load of the entire tumor mass. Its volume and composition are partly governed by the desmoplastic response driven from fibrotic cytokines (i.e. TGF-beta, PDGF, FGF2), released by the cancer tissue. For instance, the increased expression of TGF-beta in pancreatic cancer correlates with the desmoplastic response and the rate of disease progression. Platelet-derived growth factor (PDGF) is another growth factor, secreted by cancer cells, which also correlates with cancer progression. However, cancer cells do not often express PDGFRs, indicating that the growth factors do not elicit their actions via autocrine mechanisms, but involve fibroblasts and endothelial cells. In contrast to TGF-beta, PDGF seems to mostly induce proliferation of CAFs, rather than the initiation of the transdifferentiation program. Like PDGF, the basic fibroblast growth factor (FGF2), a member of the FGF family, may stimulate the proliferation of fibroblasts, but not the expression of alpha-smooth muscle actin (alpha-SMA). Despite the fact that FGF2 has more prominent role in tumor-associated angiogenesis, its direct implication in tissue fibrosis has been broadly demonstrated.

Upon their recruitment in the desmoplastic interface, CAFs elicit an altered gene- and protein-expression machinery, compared to their normal progenitors. Unfortunately, it seems that CAFs present with phenomenal phenotypic plasticity and diversity in the tumor microenvironment, which could be attributed to many possible reasons (refer to chapter 1.1.2.4 for more details), making it rather impossible to render unique and specific molecular markers for their identification. However, it is commonly accepted that most CAF subsets tend to abundantly express alpha-smooth muscle actin (alpha-SMA) and fibroblast activation protein (FAP), which are both, generally, not found in NFs. Numerous other markers, such as desmin, vimentin, discoidin-domain receptor-2 and integrin-alpha1/beta1, have been proposed as potential CAF markers. These have failed to be established as robust ones, either because they are expressed by specific CAF subsets, or because they are opulently expressed by numerous other cell types in the tumor microenvironment, and even by NFs themselves.
FIGURE 1.3. Normal and activated fibroblasts found at wound-healing and tumor sites. (A) Normal fibroblasts (NFs). NFs are instilled in the fibrillar extracellular matrix (ECM) of connective tissue, which consists largely of type I collagen and fibronectin. NFs interact with their surrounding microenvironment through integrins such as the alpha-1/beta-1 integrin. Typically though, NFs are considered as being in quiescence and appear as fusiform cells with a prominent actin cytoskeleton and vimentin intermediate filaments. (B) Activated fibroblasts or myofibroblasts or cancer-associated fibroblasts (CAFs). Fibroblasts can acquire an activated phenotype, which is associated with an increased proliferative activity and enhanced secretion of ECM proteins such as the ones shown above, and characterized the desmoplastic stroma found in most solid carcinomas. CAFs are often characterized as expressing alpha-smooth-muscle actin (alpha-SMA), among other markers. [Illustration adapted from: Kalluri and Zeisberg (2006)]
1.1.2.3. The recruitment of peritumoral CAFs in the tumor-host cell interface

CAFs participate in the construction of a tumor-associated lesion known as “desmoplasia” or “desmoplastic reaction”\(^8\). Desmoplasia is predominantly noted in the cancer invasive margins, whereby local invasion and tumor progression is a dynamic and reciprocal process\(^11\). Because cancer cells lining the invasive margins are directly juxtaposed to elements of the desmoplastic stroma (i.e. ECM components and CAFs themselves), it is strongly suggested that the structure and paracrine support of CAFs may facilitate aspects of tumor progression, through the deployment of altered protein expression machinery elicited in CAFs\(^11\). Although the literature in this field is emerging, there is a vast plethora of studies that attempt to decipher the paracrine elements, instilled in desmoplastic interfaces by CAFs\(^53,11\).

Here, we will only enumerate a few of the most important features of this *de novo* expressed protein machinery, which collectively demonstrate that CAFs induce: (A) an altered ECM structure, as compared to the one observed in normal stromata, which provides oncogenic signals to the tumor cells, i.e. the aberrant expression of tenascin-C (TNC)\(^11\) and matrix metalloproteinases, like, for example, the gelatinases MMP2 and MMP9, and stromelysin-1 (MMP3)\(^11\), (B) increased expression of growth factors and cytokines, such as insulin-like growth factor-1 (IGF1) and hepatocyte growth factor (HGF) that directly promote tumor cell survival and motility, respectively\(^11\)\(-\)\(^12\), (C) inflammatory responses by secreting chemotactic and proinflammatory agents, like, for example, IL-1\(\beta\) and tumor necrosis factor-alpha (TNF-alpha)\(^12\), and finally (D) angiogenesis by interactions with the local microvasculature, by aberrantly secreting proangiogenic VEGF and stromal-derived growth factor-1 (SDF-1)\(^19\),\(^12\). Overall, CAFs of the desmoplastic interface may initiate and maintain paracrine communication with almost all types of stromal cells present in the tumor microenvironment, besides their direct impact on the tumor cells lining the invasive margins (Fig. 1.4.).

In addition to their paracrine signaling effects, CAFs appear to exert a direct physical impact on tumor tissues\(^19\), resulting in increased peritumoral ECM stiffness and consequently mechanical stress. This may affect the malignant phenotypes and the metastatic behavior of the cancer cells\(^20\). This dual paracrine/mechanical viewpoint of the metastatic cascade will be thoroughly explored in chapter 1.1.3.
FIGURE 1.4. The paracrine milieu of pleiotropic CAF functions in the tumor microenvironment. CAFs may communicate in a paracrine aspect with epithelial cells, endothelial cells, pericytes and inflammatory cells besides cancer cells of the invasive margins, through the secretion of growth factors and chemokines. A few of the many instances, also described in the text, are listed. In brief, the deposition of ECM components like collagen types I and III and TNC that provides oncogenic signals is shown. CAFs additionally mediate inflammatory responses through secretion of chemokines, for example, monocyte chemotactic protein-1 (MCP1) and interleukins, such as IL-1b. CAFs also interact with components of the cancer-associated endothelium, especially by secreting various MMPs, SDF1 and VEGF. CAFs also provide oncogenic signals, such as TGF-beta and HGF, to resident epithelia, to directly stimulate cancer-cell proliferation and invasion. [Illustration adapted from: Kalluri and Zeisberg (2006) 53]
1.1.2.4. CAF subsets and speculations on the observed diversity

CAFs found in different types or subtypes of cancer, or even within the same cancer may share different gene and protein expression signatures. An exaggerated paradigm demonstrating the extensive stromal diversity comes from a recent approach to create molecular subtypes in breast cancer that are dependent on stromal genetic signatures and which could accurately predict prognostic outcome, instead of the more traditional molecular subtypes that are based on receptor expression in breast cancer cells. Interestingly, different CAF subtypes may also elicit diverse functional properties, an observation that makes the elucidation of this process quite important. So far, there have been two possible hypotheses, attempting to provide insight on the phenotypic diversity of CAFs.

The Memory-from-Progenitor hypothesis: CAFs may be derived from a wide variety of progenitor cells beyond the quiescent fibroblastic population of the adjacent stroma. A considerable subset may actually originate from mesenchymal stem cells or bone marrow-derived circulating cells and myeloid precursors, which are able to localize in peritumoral regions, exploit the available repertoire of signaling factors in the invasive margins and subsequently transdifferentiate into CAFs, and initiate proliferation. Some lines of evidence demonstrate the presence of a context-dependent phenotypic switching in endothelial cells, as a CAF source. For instance, TGF-beta-2 signaling in endothelial cells is shown to induce endothelial-to-mesenchymal transition (EndMT), a biological reprogramming that may generate a very unique CAF subset. Indeed, a significant proportion (up to 40%) of CAFs may share endothelial markers such as platelet/endothelial cell adhesion molecule (or cluster of differentiation-31) (PECAM/CD31), which implies that they originate from an endothelial subpopulation. Furthermore, a special case of cancer cell EMT, may actually lead to the formation of another unique CAF subset. For instance, breast cancer cells may typically undergo EMT that transforms them into myoepithelial cells and subsequent transdifferentiation into CAFs. This mechanism may also explain why spontaneous genetic studies identify cancer cell-related mutations (i.e. p53 mutations) in various tumor-associated stromal cells, which appear to belong to a non-malignant population. Finally, it should be mentioned that certain tissue-specific stromal cells may also provide an adequate pool for CAF subsets through specific signaling platforms, as is the case of the adipocyte-derived CAFs in the breast tumor microenvironment. Given the plasticity described, it is plausible to assume that each CAF subset may retain molecular memories from the respective progenitors, and therefore their gene- and protein-expression machineries may significantly differ between each other.

The Tumor-Heterogeneity hypothesis: Intratumoral cancer cell heterogeneity is now considered as a well-established concept, and implicates various cell subpopulations (known as clones) in tumorigenesis,
which may expand through independent-to-each-other molecular pathways leading to differences in drug resistance. Also, certain lines of evidence allege the possibility that these clones cooperate with paracrine mechanisms for the progression of the disease. Interestingly, our group has noticed different CAF phenotype/behavior in vitro whenever NFs were stimulated with soluble factors derived from different colon cancer cell lines (i.e. SW480/SW620 as opposed to HT29). In particular, HT29-stimulated CAFs were able to adopt collective configuration (i.e. syncytial) whenever migrating, through an aberrant over-expression of claudin-11 (CLDN11) and occludin (OCLN), two proteins of the tight-junction apparatus. On the other side, SW480/SW620-stimulated CAFs migrated in a non-collective, individualized pattern, by retaining their original spindle-like shape. Despite the fact that these cell lines were derived from different patients or from the same patient at different stages of the disease, this experiment suggests that normal quiescent fibroblasts have the capability of generating two distinct subsets of CAFs, depending on the context they are triggered with. This theory, therefore, reasons that tumor subpopulations may be able to alter the properties and generate different subsets of recruited stromal cells, including CAFs, with ultimate purpose of creating a supportive microenvironment for their own needs that may vary among the various clones.

Since there is strong evidence supporting either of the two models, we conclude that the observed stromal heterogeneity is probably a byproduct of both processes. The possibility that this intermixed CAF population found in the desmoplastic interface principally occurs due to a synergistic effect of the two mechanisms or these two are working in a non-redundant fashion to produce multiple CAF subsets, is yet to be elucidated.

1.1.3. Interdigital Migration: A CAF-Mediated Viewpoint of the Metastatic Cascade

During collective cell migration, the border (i.e. “tip”) cells, belonging to multicellular aggregates of either tightly or loosely-associated cell-groups, are principally responsible for driving migration of the entire cohort, and, quite interestingly, cells positioned in the core of the collectives might not even retain a migratory phenotype or genetic properties of cell motility. Such is also suggested to be the pattern of cell migration in most solid cancers (Fig. 1.5.). Having undergone a sequential series of genetic alterations, cancer cells eventually breach the underlying basement membrane, escaping from the in-situ state, and excavate multiple pathways towards lymphatic or blood vessels. Cancer cells, originating from mucosal epithelia, do not migrate alone in the underlying submucosal cell layers, but rather in form of cohort/collective configuration, especially during the early events of the metastatic cascade.
In such collectives (Fig. 1.5.), there is a subpopulation of cancer cells lining the interface, also known as the “cancer invasion front” (CIF), which is currently speculated as the driver of cancer cell migration\textsuperscript{110, 145}. On the inner side of CIFs, the so-called “tumor-core”, there are heterogeneous subpopulations of cancer cells, which may be entirely unaffected from paracrine signaling with the cancer-associated stroma. Such cells may also retain more epithelial-like phenotype and characteristics than those in CIFs and may also express homotypic cell adhesions through functional adherens and tight junctions, which provide strong anchoring milieu with the corresponding CIF\textsuperscript{110}. However, cancer cells belonging to the CIF exert a phenotypic behavior that is constantly under the adjacent microenvironmental pressure, which is responsible for driving localized invasion\textsuperscript{110, 145}. The most recent paradigm, attempting to explain the pattern of cancer cell migration is that of the “invasive growth”\textsuperscript{146}, a special form of collective migration where tumor core cells may excessively proliferate, causing constant mechanical pressure in the CIF cells. The latter cells develop a gene- and protein-expression machinery, which may be context-dependent as implied already, to allow them to penetrate further into the stroma\textsuperscript{64,144,146}.

However, certain groups\textsuperscript{147-152}, including ours\textsuperscript{127}, have been investigating the concept that the stromal compartment in these CIFs is not quiescent as implied from this paradigm, but develops an enhanced migratory capability and may exert a coordinated, syncytial-type of activity against the cancer cells themselves. By collectively taking into consideration these observations, here, we propose a detailed working model of metastatic growth progression, based on both the paracrine and the mechanical impact of the CAF cohort at the tumor-host cell interface (Fig. 1.6.; middle and right columns)\textsuperscript{64}. In this model, the interaction between cancer cells and normal host stroma results in the induction of desmoplastic reaction, characterized by the emergence of a responsive myofibroblastic tissue, the CAF cohort. The CAF cohort acts both as a myofibroblastic-signal source niche and as a migratory cohesive unit, interacting mechanically with the epithelial compartment (Fig. 1.6.; middle and right columns). At the tumor invasive front, it creates a complex and dynamic framework, in which stromal populations invade the epithelial ones and vice-versa (Fig 1.6.; left column). The invasive properties of each population, together with the mechanical stress of the microenvironment may actually direct (and to some extent enforce) these populations to migrate against each other in an interdigital pattern (Fig. 1.6.; left column).
FIGURE 1.5. The tumor invasion front. Observations on longitudinal sections (top-left corner) of cancers penetrating the host stroma show finger-like configuration of cancer spread. After penetrating the basement membrane, epithelial cancers migrate toward the underlying lymphatic or blood vessels, patterned in a branching morphology. In such areas within the stroma, transverse section (yellow line) may reveal patterns of cancer cell "cohorts" (top-right corner). In this cartoon schematic, cancer cell cohorts are depicted with black, blood vessels with red roundish shapes and stromal cells (CAFs) are depicted with the background green color. A magnification of the tumor-host cell interface area reveals two clearly distinguished subpopulations, the cancer population, and the CAFs or myofibroblasts. The magnified area is depicted through both a cartoon (left schematic) and a histologic figure (right schematic), obtained from our archive. This interaction area is described as the "tumor invasion front" and it is characterized by "desmoplasia", a histopathological lesion defined as the peritumoral accumulation of CAFs with parallel deposition of ECM components.
FIGURE 1.6. "Interdigital migration," a working model of CAF-directed metastasis. The cartoon illustrates 4 (I–IV) sequential steps of the proposed model of interdigital CAF-directed metastasis and briefly shows how the myofibroblastic or CAF cohort invades the cancer population and vice-versa. The illustration is followed by the reader vertically in each step (I–V), whereby description for the cell-population (left column), mechanical (middle column), and paracrine signaling (right column) dynamics are provided for each specific snapshot during the metastatic process. The first column shows the population movements: cancer cells, blue roundish; CAFs, green elongated; ECM, brown curvy lines. The second column represents the exact same snapshot as the first column: cancer cells, blue; CAFs, green; arrows within the snapshot show mechanical tensile forces inflicted by one population over the other; the thickness of the gray line illustrated over the snapshot and shown by black arrow is relative to the mechanical pressure on the cancer population at the level of the red arrow. Therefore, before invasion both subpopulations rest in quiescence (step I), so the thickness of the gray line is very small. After the first CAF digits invade the cancer population (steps II and III), the thickness of the gray line significantly and progressively grows. This mechanical force allows the cancer cell cohort to invade the CAF cohort in an antiparallel manner, reducing the mechanical tension at the level of the invasion front (step IV), so the thickness of the gray line in the last snapshot is again reduced. The third column represents the exact same snapshot as the first column: cancer cells, blue scale; CAFs, green. The blue scale in the cancerous population is proportional to the differentiation status of the cells in each step (I–V) of metastasis, the darker the blue, the less differentiated the cancer cells. Note that the paracrine impact of the CAF cohort translocates the signaling gradients (rectangular box) to retain the undifferentiated state of the cancer cells lining the invasion front.
1.1.4. Tumor Budding

An exaggerated derivative of the process of local invasion may be the development of “tumor budding”, which has been described as a subset of isolated individual or clustered (<5) cancer cells that have detached from proximal invasion fronts and have further penetrated into the underlying, neighboring stroma\(^{153,154}\). The “tumor buds” or “tumor-budding cells” tend to form nest-like structures, as usually seen longitudinally under the microscope, and are almost indistinguishable from the surrounding (typically desmoplastic) stroma, since they have lost their pure epithelial phenotype and partially share mesenchymal and stem-like properties\(^{153}\).

Tumor budding at desmoplastic CIFs has been recognized as an adverse parameter and independent prognostic factor by the International Union against Cancer (UICC)\(^{155}\). High-grade tumor budding has been associated with lymph node metastasis\(^{156-158}\), distant metastasis\(^{159}\), local recurrence\(^{160}\) and correlates with the distance of tumor invasion beyond the outer border of the \textit{muscularis mucosae}\(^{161}\).

An imperative biological determinant for tumor-budding formation is the merging of contextual signals that together create motility and differentiation gradients across the cancer cell cohorts/collectives invading the cancer-associated stroma. Here, we briefly demonstrate vital evidence for the existence of such gradients, despite the fact that the contextual signals regulating them are poorly understood (and comprise a biological question that we try to address in this thesis). A prominent example for the existence of motility gradient across the cohort is fascin, a well-established marker of enhanced motility, which is strongly expressed in the periphery of tumor cohorts (i.e. CIFs), and is progressively lost towards the tumor core cells\(^{162}\). Examples for the existence of differentiation gradient across the cohort come from observations on the EMT biological program. Collectively, these studies underscore that (a) tumor core cells have not yet, (b) cancer cells lining the CIFs have only partially, and (c) tumor buds have fully undergone an EMT program [for details refer to the article by Zlobec\(^{154}\) and references therein]. As a proof-of-concept, several groups\(^{99,154}\) have demarcated stronger nuclear expression of beta-catenin and more significant loss of membranous expression of E-cadherin and OCLN in tumor budding compared to CIF and tumor core cells, thus denoting the presence of a dedifferentiation towards the desmoplastic interface. Notably, many prominent EMT-initiating transcription factors, such as Zeb and Snail, are highly expressed in tumor buds, moderately expressed in CIFs, and almost absent in tumor cores, which again suggest the function of an EMT biological cascade, in close proximity to the cancer-associated stroma\(^{163}\). The identification and characterization of precise contextual signals, responsible for mediating these gradients remain to be elucidated.
1.2. ONCOPROTEOMICS AND CANCER SECRETOMICS

1.2.1. General Notions

Novel proteomic tools and mass spectrometry play a central role in protein research, especially in the simultaneous identification, quantification and characterization of thousands of proteins, even in complex biological samples \(^{164}\). The emergence of mass spectrometry-based proteomics enabled the field of cancer research with new opportunities. The field of “oncoproteomics” deals with applications of proteomics in clinical and molecular oncology \(^{165}\).

Among the plethora of available tissues/fluids for proteomic analysis, in this PhD thesis, we focus on “cancer secretome”. Secretome analysis has only recently been established as a sub-field of oncoproteomic, but the indications thus far suggest that it is a promising source of biomarkers and therapeutic targets for various types of cancer. Secreted proteins account for approximately 10-15% of the proteins encoded by the human genome and participate in various physiological processes such as immune defense, blood coagulation, matrix remodeling and cell signaling, but also in pathological conditions including cancer angiogenesis, differentiation, invasion and metastasis \(^{166}\).

1.2.2. Protein Secretion Pathways

To better understand the nature of secretome analysis, we first provide a brief overview of protein secretion pathways, responsible for the presence of a large number of extracellular proteins in the microenvironment of normal and/or cancer cells. In eukaryotic cells, soluble proteins are secreted in the extracellular space either by exocytosis of secretory vesicles or by release of secretory/storage granules upon stimulation and activation of intracellular signaling pathways. The secreted proteins are mostly synthesized as protein precursors, which contain N-terminal signal peptides that direct them to the translocation apparatus of the endoplasmic reticulum (ER). These proteins are transported to the Golgi apparatus and subsequently to the cell surface, where they are liberated into the microenvironment by fusion of the Golgi-derived vesicles with the plasma membrane. This well-characterized protein secretion pathway has been termed as the “classical secretory pathway” \(^{167,168}\).

Other lines of evidence point out that in addition to this mechanism, proteins can be exported by various ER/Golgi-independent pathways, the so-called “non-classical secretory pathway”. At least four distinct types of non-classical exports have been distinguished over the years, all of which lack the presence of the classical signal peptide for ER/Golgi dependent protein secretion \(^{169}\). Certain proteins, such as Interleukin-
1 beta (IL-1b), are imported into intracellular vesicles, which are endosomal compartments and through a process called endosomal recycling, they are released in the extracellular space upon fusion of the endosomal vesicle with the plasma membrane. Other proteins, such as fibroblast growth factor-1 and -2 (FGF-1 and -2), reach the extracellular space by direct translocations across the plasma membrane using distinct transport systems. Another proposed mechanism of non-classical protein secretion involves the direct translocation of the protein to the extracellular space, but it requires that the protein is membrane-anchored through dual acylation in the N-terminus, and a flip-flop mechanism mediates the secretion. Finally, proteins can also be secreted through exosomes; these vesicles originate from the internalization of activated receptors along with scaffold proteins surrounding them, followed by traffic through early endosomes. These receptors are further internalized within the endosome, forming the late endosome, known as “multivesicular body” (MVB). Internalized receptors within the late endosomes are referred to as “intralumenal vesicles” (ILVs) when they are present within the MVBs, but are referred to as “exosomes” upon fusion of the MVB with the plasma membrane and subsequent secretion.

1.2.3. Defining the Cancer Secretome

The term “secretome” was originally adapted by Tjalsma et al. and Antelmann et al. as a concept providing an integrated and global view of the protein secretion by considering both secretion systems and their protein substrates. A sequential term has been later adapted, and this was the “exoproteome”, which actually refers to the entire extracellular proteome, and particularly suggests that this subset of proteins is localized in the extracellular milieu and has strong potential of getting released into the circulation.

Inherent to the description of the various secretory pathways above, the cancer secretome has been described as including the ECM components and all the proteins that are released from a given type of cancer cells, such as growth factors, cytokines, adhesion molecules, shed receptors and proteases, and reflects the functionality of this cell type at a given time point. Therefore, the cancer secretome includes proteins released from cancer cells, either with classical or non-classical secretory pathways, and corresponds to an important class of proteins that can act both locally and systemically. Therefore, the cancer secretome includes all proteins that can be identified in the interstitial fluid of the tumor mass in vivo, however it is better conceptualized as the group of proteins identified with mass spectrometry (MS) in cancer cell line conditioned media (CM) in in vitro studies.

As earlier noted, primary tumors are composed of not only cancerous cells, but also of a wide diversity of stromal cells, which are recruited as active collaborators, facilitating the development and progression of malignancy. Out of these heterotypic interactions, a great variety of proteins, including growth factors,
enzymes such as proteases, smaller protein molecules like chemokines and cytokines, as well as many other proteins are constantly released from all participating cells and act upon others in an autocrine or paracrine fashion, resulting in the acquisition of a favorable milieu for the progression of the malignancy (Fig. 1.7.). Cancer cell-secreted proteins alone comprise only a subset of the overall microenvironmental proteins in an in vivo context. Therefore, the investigation of the recruited stromal secretome constitutes a critical strategy for the identification of novel biomarkers or key regulators of carcinogenesis that could be assigned a therapeutic potential.

In this thesis, we propose for the first time a wider definition for the already-established term “cancer secretome”, in which it additionally includes the extracellular proteome derived from stromal cells as a result of tumor-host cell interactions, apart from the cancer cell-secreted factors. To better demonstrate this, we provide an illustration of the “heterotypic cancer secretome” below, in which all secreted proteins are assigned a possible origin from many types of cells within the tumor microenvironment (Fig. 1.7.).
FIGURE 1.7. Heterotypic overview of the cancer secretome. All the microenvironmental, secreted proteins may originate either from cancer cells or from associated stromal cells and their secretion may be triggered by paracrine or autocrine actions between/among them. Proteomic approaches to capture the tumor microenvironment should focus on identifying proteins secreted by all associated cells, not just the cancer cells. Arrows initiating from cells and pointing in molecules represent secretion; the opposite represents the paracrine or autocrine action of the secreted molecules on the cell types.
1.2.4. Sources of Cancer Secretomes

Two prominent sources have been utilized in cancer secretome studies: cancer cell line supernatants and proximal biological fluids. A major opposition to tissue/cell culture is the inability to fully replicate the complexity of the tumor microenvironment in vivo. For instance, changes in protein expression may occur because of cell culture stress instead of having certain in vivo relevance. In one study, the authors found significant changes in protein expression, even after short-term culturing of low-grade superficial bladder transitional cell carcinomas in vitro. In another study, the authors noticed significant changes in protein expression between micro-dissected prostate cancer cells and cell lines developed from the same patient, further demonstrating that culture stress may affect the differential protein potential. Although certain such differences exist, it is generally accepted that the genomic and transcriptional profiles of cancer cell lines have been adequately recapitulating the most leading features of primary tumors. The identification of many known cancer biomarkers in the conditioned media of cancer cell lines supports that it is a representative source to mine, despite the fact that it lacks the contribution of the tumor microenvironment.

Mining the secretome from cancer tissues would constitute an appealing alternative approach to study proteins produced in vivo by the tumor, but it has been rather under-studied, due to evident technical challenges. Another alternative would be fluids that are proximal to tumors, which are shown to frequently contain cancer cells, in addition to numerous soluble growth factors released by cancer cells and the tumor microenvironment. Many proximal fluids can be obtained with minimally invasive procedures and in large amounts (e.g. ascites fluid from ovarian cancer patients). However, the procedures to obtain such fluids need to be standardized. In addition, since samples are collected from different individuals, the variability caused by behavioral, environmental and genetic differences is unavoidable. Furthermore, contamination by highly abundant serum proteins can increase sample complexity and complicate data interpretation.

On the contrary, the cancer cell line CM contains secreted or shed proteins, released through classical and non-classical secretion pathways, and its limited complexity compared to serum, proximal fluids and cancer tissues enhances identification of low abundance proteins. Moreover, as in any in vitro system, experimental conditions can be highly controlled allowing reproducible and quantifiable results. Furthermore, large numbers of cell lines representing various stages and histotypes of a given cancer are readily available; the US National Cancer Institute (NCI)-60 human tumor cell line anticancer drug screen, developed approximately two decades ago as an in vitro drug discovery tool, represents the most notable example of such availability.
However, no single cell line can recapitulate the heterogeneity of human tumors. Cell lines, for the most part, are deficient from contributions in the host-tumor microenvironment. In addition, genotypic and phenotypic alterations accumulating over time may give rise to distinct subpopulations in the same cell line. The presence of serum is important for cell survival and growth under in vitro conditions and frequently conditioned media are supplemented by an exogenous source (e.g. fetal bovine serum; FBS). At the same time, serum starvation has been shown to affect cell survival, proliferation, protein production and secretion patterns \(^{191-195}\). However, in the majority of secretome analysis studies, cells are grown in serum-free media. This approach reduces both sample complexity caused by high protein content of FBS and sample contamination with orthologous proteins that may share amino acid sequences with the proteins of interest. One alternative approach was proposed by Colzani \textit{et al.} and involved the supplementation of isotopically labeled amino acids in FBS-containing CM, resulting in the isotopic labeling of proteins that originate from cells \(^{196}\). Although in this study, labeling made it possible to distinguish between cell-derived and bovine proteins, additional steps in sample preparation were required to deal with increased sample complexity. A more user-friendly approach, which has also been tested in our laboratory, would be to adapt the cells to serum-free media by gradually reducing the percentage of serum in the CM prior to proteomic analysis.

As noted above, the major obstacle in the study of actively secreted proteins in the CM is the passive release of proteins into the media caused by cell death. Given that secreted proteins are of low abundance, they can be easily “masked” by highly-abundant intracellular proteins. For that reason, frequently, cells are incubated in serum-free media only for a small period of time such as 24h \(^{197-199}\). However, the amount of total protein secreted by the cells in 24h is rather small. In our laboratory, we have established an optimization procedure to maximize protein secretion and minimize cell death \(^{200}\). In our workflow, multiple seeding densities and incubation periods are tested and levels of total protein, cell death and protein secretion are monitored for all different conditions. Based on these parameters, optimum conditions are selected. More sophisticated approaches, such as hollow fiber culture systems and nanozeolite-driven enrichment of secretory proteins have occasionally been reported \(^{201,202}\), but these may be particularly time and resource consuming.

1.2.5. \textit{In Silico} Purification of the Cancer Secretome

Not all proteins identified in the CM or biological fluids during secretome analysis can be considered \textit{per se} as actively secreted proteins, and as such, many may be contaminants resulting from cell death. This issue may be addressed through deployment of \textit{in silico} filtering for removal of the intracellular
contaminants, despite the fact that standardized methods to achieve so have been lacking. Of note, certain bioinformatics tools can distinguish between secreted proteins and intracellular contaminants.

One of the most widely-used principles to filter-out intracellular contaminants is the systemic mining of publicly available databases that have information on protein subcellular localization. One such prominent database is Gene Ontology (GO) (available at http://www.godatabase.org/dev); however, an advanced understanding of GO structure is critical to interpret the data correctly. The NCBI PubMed (http://www.ncbi.nlm.nih.gov), Swiss-Prot/TrEMBL (http://www.expasy.org), and Bioinformatic Harvester EMBL (http://harvester.embl.de/) also comprise databases with protein cellular localization information, based on literature findings. Finally, Human Proteinpedia comprises a community portal that acts as a reservoir of human protein data and Human Protein Reference Database (HPRD), and is used to integrate data deposited in Human Proteinpedia.

Another principle for filtering-out intracellular contaminants is the in-silico prediction of proteins that are secreted, based on their primary sequence. This is especially useful in proteomic datasets, in which experimental data on protein secretion are not available. For classical secretion, these algorithms screen a target sequence in search of N-terminal signal sequence or a signal sequence cleavage site. A few of these are SignalP (http://www.cbs.dtu.dk/services/SignalP) 205, web-based secreted protein database (SPD) (http://spd.cbi.pku.edu.cn) 206 and Signal Peptide Prediction (SIG-Pred) (http://www.bioinformatics.leeds.ac.uk/prot_analysis/Signal.html). Combination of multiple methods may increase predictive accuracy. For non-classical secretion, algorithms such as SecretomeP (http://www.cbs.dtu.dk/services/SecretomeP) 208 are used. In addition, given the recent observations on exosomal proteomics, an independent database of proteins secreted through these endocytic-like vesicles, named ExoCarta, has been generated (http://exocarta.ludwig.edu.au/index.html) 209. Finally, it is also possible that proteins located on the plasma membrane are shed and released to the extracellular space. Some algorithms that predict such sequences are TransMembrane prediction with Hidden Markov Models (TMHMM) (http://www.cbs.dtu.dk/services/TMHMM) 210, and prediction of Transmembrane Regions & Orientation (TM-pred) (http://www.ch.embnet.org/software/TMPRED) 211.

A summary of this section is shown in Figure 1.8., which schematically illustrates all protein secretion pathways, and is also accompanied by the proposed prediction algorithms (blue squares).
FIGURE 1.8. Bioinformatics tools for prediction of protein secretion pathways. The classical protein secretory pathway is ER/Golgi-dependent and involves the presence of the signal peptide that directs translocation of these proteins to the ER. Protcomb, SignalP, SPD and Sig-Pred are some of the widely used programs for prediction of proteins secreted through the classical secretory pathway. The non-classical protein secretion pathway is ER/Golgi-independent and is associated with absence of signal peptide. SecretomeP has been used for prediction of proteins secreted through non-classical secretion pathways. Proteolytic events in the extracellular space might also result in shedding of membrane-bound proteins/particles. Although this is not a protein secretion pathway, but an extracellular proteolytic event, software, such as TMHMM and TMPred is being used for the prediction of membrane and membrane-bound proteins. Finally, a database of exosome-secreted proteins, called ExoCarta has been recently generated as a distinct database the proteins secreted as such. ER, endoplasmic reticulum; ILVs, intralumenal vesicles; MVBs, multivesicular bodies; SPD, secreted protein database; SIG-Pred, signal peptide prediction; TMHMM, transmembrane prediction with hidden Markov models; TMPred, prediction of transmembrane regions and orientation. Arrows indicate protein secretion or vesicle processing/movement. Blue boxes indicate the suggested software.
1.2.6. The Heterotypic Overview of the Cancer Secretome

As demonstrated earlier, it is now firmly believed that contextual signals around the tumor may not only be important, but also necessary for regulating the malignant phenotype and behavior of cancer cells, particularly those lining the tumor invasion front \(^4, \^{110}\). This anatomical site may also serve as the major driver of metastasis \(^{110, \^{144}}\), viewed as a sequential process involving the detachment of individual cancer cells from the cohort (the so-called tumor buds) and their subsequent invasion and migration in the neighboring desmoplastic stroma \(^86, \^{154}\). It is strongly suggested that an EMT event may be sufficient to provide properties of cell stemness in the tumor budding subpopulation and efficiently mediate and terminate this process \(^99, \^{212-215}\).

Given the above, it seems that the field of cancer secretomics could play a substantial role in properly defining these microenvironmental signals that together form the context of the invasive-edge- and tumor-budding-subpopulations and possibly also regulate EMT. As such, we have provided a new definition for the Cancer Secretome in section 1.3.3., by combining the collective knowledge from a plethora of proteomic studies \(^{166}\) and references therein] with the heterotypic viewpoint, as described in the landmark article “Hallmarks of Cancer” \(^2\).

Here, we will focus on three distinct enabling characteristics of every neoplastic microenvironment (inflammation, desmoplasia and angiogenesis), each one bearing unique stromal participants. In the course of describing these, an effort will be made to explain how cancer secretomics may play a fundamental role in linking the contextual signals derived from the tumor-host cell interface with a variety of cell-biological programs (i.e. EMT), an aggressive phenotype, localized invasion/migration (i.e. tumor-budding formation) and, in general, the malignant behavior of the cancer cells at the level of the invasion front.

Inflammation has been an enabling characteristic of the tumor microenvironment and is now believed to possess an unanticipated, paradoxical effect of enhancing tumorigenesis and progression, in effect helping incipient neoplasias to acquire hallmark capabilities \(^38, \^{41, \^{48, \^{216, \^{217}}\). Most immune cell lineages are now seen as significant components of neoplastic tissues, however, we will only focus on the tumor-associated macrophages (TAMs). TAMs are derived from monocytes and are recruited by monocytic protein chemokines, secreted by the cancer cells in various contexts \(^35, \^{48, \^{52}\). Upon differentiation, TAMs secrete a considerable number of angiogenic and lymphangiogenic growth factors, cytokines and proteases, all of which are mediators of neoplastic development and progression \(^35, \^{52, \^{218-221}\). Therefore, the interactions between TAMs and cancer cells have been investigated for a long time, but only recently, proteomic technologies have been deployed for studying the altered secretion profiles of these cells. In one such
study, the authors performed secretome analysis using LC-MS/MS on supernatants from a normal monocyic/macrophage cell line, buffy coat monocytes, as well as purified, in vitro-cultured TAMs, isolated from ovarian cancer ascitic fluid and they noticed the de novo secretion of 14-3-3 zeta protein in cancer-associated macrophages \(^{222}\). Given the previously documented role of 14-3-3 zeta protein as adaptor protein in intracellular signaling pathways \(^{223,225}\), and more recent evidence that this protein can also be secreted in the extracellular space by monocytes/macrophages infected with HIV-1 virus \(^{226}\), the authors speculated that TAM-secreted 14-3-3 zeta protein may promote neoplastic progression of epithelial ovarian cancer, under conditions which promote its secretion. Remarkably, this was the first mass spectrometry-based study that demonstrated a novel mechanism of neoplastic progression, mediated by cancer cell-TAM interactions, using secretome analysis.

As earlier described, the recruited fibroblasts (CAFs) around cancer cells expand across most solid carcinomas and enhance tumor phenotypes, notably cancer proliferation angiogenesis, and invasion and metastasis. The full spectrum of functions contributed by CAFs depends on the increasingly diverse repertoire of signaling factors that have been discovered over the years \(^{5,47,53-58,63}\). As such, we expect that cancer secretomics may offer a fruitful platform for fully elucidating these molecular networks and for advancing our understanding on tumor pathogenesis. In one proteomic study, the authors sought to investigate the mammary CAF secretome, so they induced the myofibroblastic phenotype by generating CAV1 (-/-) fibroblasts, based on the hypothesis that since caveolin-1 inhibits TGF-beta signaling, then CAV1 (-/-) fibroblasts could maintain a constantly active TGF-beta pathway, which is known to trigger the induction of CAFs. Secretome analysis of CAV1 (-/-) fibroblasts indicated the secretion of factors associated with the myofibroblastic phenotype (e.g. Colla1, Colla2 and SPARC), verifying the initial hypothesis \(^{227}\). In another study \(^{228}\), the authors cultured melanoma and associated stromal cells, including melanoma-associated fibroblasts and normal skin fibroblasts and performed mass spectrometric analysis in cell lysates and supernatants using LC-MS/MS. Their analysis showed many melanoma-specific secreted proteins (lumican, Pmel 17), as well as proteins secreted by normal (extracellular matrix proteins) or melanoma-associated fibroblasts (neuropillin, stannioiclin-1, periostin). This strategy provided novel insights into secreted proteins, not been previously identified in melanoma, such as glutathione peroxidase-5 (GPX5) \(^{228}\).

Like normal tissues, tumors require sustenance in the form of nutrients and oxygen, as well as the ability to evacuate metabolic wastes and carbon dioxide. The tumor-associated neovasculature, generated by the process of angiogenesis, is a major phenomenon of every tumor microenvironment and addresses these needs. A compelling body of evidence indicates that the angiogenic switch is governed by countervailing factors that either induce or oppose angiogenesis. Some of these angiogenic regulators are secreted
signaling factors that bind to stimulatory or inhibitory cell-surface receptors, displayed by vascular endothelial cells \(^6, 7, 9, 17, 229, 230\). Therefore, model systems of hypoxic and angiogenesis-regulating microenvironments are an attractive field of cancer secretomics for the systemic exploration of these molecular events. For instance, in a co-culture model of murine lung cancer along with stromal cells, including murine endothelial cells, a wide multitude of cytokines were found to be over-expressed in the co-cultures compared to the monocultures, when quantified with SILAC \(^{231}\). This analysis pointed out that endothelial cells are essential and able to stimulate \textit{in vitro} and probably \textit{in vivo} the production of various soluble factors that assist in tumor development and progression.

In conclusion, cancer secretome analysis represents a nurturing but rapidly-evolving tool in the field of tumor-host cell interactions. Given its documented potential, we anticipate that it will be certainly brought to the frontiers of cancer investigation in the future.
1.3. THE BONE MORPHOGENETIC PROTEIN (BMP) PATHWAY

1.3.1. The BMP Family of Proteins: Ligands and Their Receptors

Bone morphogenetic proteins (BMPs) were originally described to act as bone growth factors. They are a group of cytokines, with an ability to regulate formation of bone and cartilage, but currently demonstrated to exert a wide range of morphogenetic activity that is both tissue- and context-dependent. The BMP family of cytokines comprises over 20 different ligands that belong to the Transforming Growth Factor-beta (TGF-beta) superfamily.

BMPs are expressed as large precursor proteins. The raw molecule carries an N-terminal signal peptide, which directs the protein to the secretory pathway, a pro-domain that ensures proper folding, and the C-terminal mature peptide. Each monomer is stabilized by three intramolecular disulfide bonds, formed between six highly conserved cysteines, a structure known as cystine-knot motif. The active signaling molecule is typically formed through homodimerization. Dimers are covalently linked via a disulfide bond, requiring a seventh conserved cysteine within each monomer. There are a few exceptions, such as growth differentiation factor-9 (GDF9), -9B (GDF9B), also known as BMP15, and -3 (GDF3), which lack the seventh cysteine, but nevertheless appear to be biologically active. Furthermore, heterodimerization has also been observed in vivo and in vitro, as in the case of BMP2/5, BMP2/6, BMP2/7, and BMP2b/7, and quite interestingly, they were described as more potent activators of BMP signaling than their corresponding homodimers. Cleavage of the pro-protein by serine endoproteases within the trans-Golgi network produces the mature protein, which is subsequently secreted. However, uncleaved pro-proteins have also been observed, despite being unable to bind to the receptor and initiate downstream signaling events. Remarkably, the GDF-associated serum protein-1, a BMP antagonist that inhibits GDF8 and BMP11, can bind to the GDF8 propeptide, which further supports the notion that such uncleaved BMPs may have functional/physiological roles in certain contexts, but these remain to be elucidated.

BMPs bind to heteromeric receptor complexes composed of type I and type II transmembrane serine/threonine kinase receptors. A total of six different receptors have been identified and known to covalently bind to BMPs. Type I receptors include Activin receptor type IA (ActRIA or Alk2), BMP receptor type IA (BMPRIA or Alk3) and BMP receptor type IB (BMPRIB or Alk6). These receptors form heteromeric complexes with the following type II receptors: BMP receptor type II (BMPRII), Activin receptors type IIA and IIB (ActRIIA and ActRIIB). The concerted action of both receptor types is necessary for BMPs to activate specific signaling pathways. The receptors contain an extracellular ligand
binding domain and an intracellular serine/threonine kinase domain. The type I receptor carries two additional motives, a glycine/serine-rich region preceding the kinase domain (GS-box) and a short region of eight amino acids, termed L45 loop, within its kinase domain. The type II receptor kinase is constitutively active. Activation of type I receptor kinase requires ligand binding, ligand–receptor oligomerization, and transphosphorylation of its GS-box via the type II receptor. The receptors are highly homologous, but non-redundant as they may participate in distinct developmental processes, and can activate both Smad and non-Smad signaling, depending on the context.

1.3.2. The BMP Signal Transduction Pathway

The BMP signal transduction pathway is a highly-complicated pathway, comprised of multiple ligands/antagonists, receptor types, intracellular mediators, intracellular regulatory elements and machineries, transcription factors, co-activators/co-repressors, as well as components of signaling cross-talk and bypass pathways. Here, we will attempt to provide the basic principles of how the BMP signal transduction pathway is organized.

BMP ligands can bind to type I and type II receptors with differing affinities. For example, BMP2 and BMP4 demonstrate high affinity for BMPRIA and BMPRIB and very low affinity for BMPRII. On the other side, BMP7 preferentially binds to ActRIIA and ActRIIB, while its affinity for ActRIA is more limited. Such differences may cause a context-dependent oligomerization of the corresponding receptors, which, in turn, results in a profound diversity in the activation of signaling cascades. However, the oligomerization process is not only determined by biochemical affinities between ligands and receptors, but also by their expression levels, as well as the presence or absence of preformed type I/type II receptor complexes on the cell surface. Several studies demonstrate that BMPs can activate distinct pathways by signaling through different receptor complexes. For instance, BMP2 is capable of oligomerizing the signaling complex in at least two different ways (Fig. 1.9.). It can either bind to its high affinity receptor BMPRIA, upon which BMPRII is recruited into the complex as a secondary event, or it can directly bind to a preformed complex (PFC) of BMPRIA and BMPRII. The former oligomerization mode is termed BMP-induced signaling complex (BISC) and results in internalization via caveolae and activation of non-Smad pathways, such as the mitogen-activated protein kinase (MAPK) signaling. The latter oligomerization mode, i.e. the PFC, may trigger Clathrin-dependent internalization and further initiate the Smad pathway. These Smad and non-Smad signaling cascades have been referred to, as “canonical” and “non-canonical” BMP pathways, respectively.
As demonstrated above, BMP ligand binding to PFCs, followed by Clathrin-mediated endocytosis has been shown to propagate Smad signaling \(^{255, 256}\). A total of three subclasses of Smad proteins have been described, which are the receptor-regulated Smads (R-Smads), the common mediator Smad (co-Smad), and the inhibitory Smads (I-Smads). R-Smads are generally prone to nucleocytoplasmic shuttling, depending on the status of a nuclear signal, present in their C-terminus. Upon phosphorylation of R-Smads (Smad1/5/8) by type I receptors, their nuclear signal is exposed and the affinity for Smad4 is increased. Consequently, a trimeric complex formation with co-Smad (Smad4) is invoked, which results in the translocation of the entire complex into the nucleus. In the nucleus, the complex assembles with multiple gene promoters to regulate target gene transcription in cooperation with an interrelated network of co-repressors and/or co-activators \(^{257-259}\). Despite the Smad pathway is regulated in multiple levels, the most well-characterized mechanism is that of the negative feedback loop, instilled from the I-Smads, namely Smad6 and Smad7 \(^{260-262}\). Major mechanisms of regulation by the I-Smads include competition with R-Smads for the activated type I receptor binding site, as well as the formation of non-functional complexes with Smad4, which reduced the pool of available co-Smads \(^{260}\), while additional mechanisms, such as co-repressor suppression- and proteasome-dependent mechanisms of BMP inhibition have also been reported \(^{263, 264}\).

Alternatively, the localization of BISCs in lipid rafts and internalization via caveosomes triggers non-Smad pathways and facilitates non-canonical signal attenuation \(^{250, 255, 265}\). Importantly, BMP was shown to induce the MAPK pathway downstream of BISC through association of BMPRI with TAK1-TAB1, and X-linked inhibitor of apoptosis protein (XIAP), leading to an activation of various kinases, such as p38, MAPK and/or c-Jun N-terminal kinases (JNK) \(^{266}\). These kinases translocate into the nucleus, where they may activate various transcription factors, for example activating transcription factor-2 (ATF2), c-Jun and c-Fos to initiate transcription of BMP target genes, including Osteopondin (OPN), alkaline phosphatase (ALP) and collagen type I \(^{267}\).

Interestingly, BMP receptors are able to control immediate early cellular processes like cytoskeletal rearrangements through non-transcriptional pathways, including activation of phosphoinositide 3-kinase (PI3K), small Rho-GTPases (i.e Cdc42) and LIM-domain kinase-1 (LIMK1) \(^{268-271}\). All of these pathways are implicated to play a major role in cell polarity and migration \(^{272}\). It is possible that non-transcriptional BMP responses depend on the recruitment of co-receptors or crosstalk to additional pathways \(^{234}\).
FIGURE 1.9. The BMP signaling cascades. (A) Signaling via BMP-induced signaling complex (BISC). BMP ligand binds to its high affinity type I receptor (BMPRI) upon which the type II receptor (BMPRII) is recruited into the complex. This oligomerization mode results in caveolae-dependent internalization and activation of non-Smad pathways. (B) Signaling via preformed complex (PFC). BMP ligand binds to a PFC of type I and type II receptors, which targets the Clathrin-dependent internalization into endosomes. This signaling route is required for progression of the Smad pathway. Crosstalk, finetuning and non-transcriptional responses are initiated at both sides. [Illustration adapted from: Sieber et al. (2009)]

Illustration adapted from: Sieber et al. (2009).
1.3.3. Regulation of the BMP pathway through BMP Antagonism

The BMP pathway may be regulated in the extracellular level through BMP antagonists/inhibitors and coreceptors, as well as in the intracellular level through a variety of signaling circuitries. The latter is beyond our scope, since we mainly focus on the extracellular proteome (i.e. exoproteome) in this thesis, and our investigations on the regulation of the BMP pathway are limited in the level of BMP antagonism. For this purpose, we suggest a comprehensive review article, which effectively summarizes the multiple levels through which intracellular (i.e. downstream) BMP signaling is regulated, either these constitute feedback loops or cross-talks with other pathways. BMP signaling can also be regulated in the extracellular level by various decoy receptors, such as BMP and activin membrane-bound protein (BAMBI) and tyrosine-protein kinase transmembrane receptor (ROR2), which usually resemble one of the two BMP receptor types, thus competing for binding with the other, and ultimately reduce the bioavailability of the active receptor complexes.

The extracellular BMP antagonists represent a number of secreted peptides, which bind BMPs with high affinity and prevent their interaction with their specific receptors. Much of the knowledge of these BMP binding proteins derives from studies conducted in Xenopus gastrula, where these antagonists may induce dorsalization of the embryos by opposing the BMP ventralizing effects. The protein sequence of the extracellular BMP antagonists is characterized by an array of evolutionarily conserved cysteine-rich (CR) domains, corporal for the formation of the cystine-knot structures. By comparing the spacing of the cysteine residues in the cystine rings, protein sequence alignment studies have categorized the extracellular BMP antagonists into the following subgroups: noggin, the chordin family, twisted gastrulation, and the CAN family.

Perhaps the most widely explored BMP antagonist is noggin (NOG), which is secreted as a glycosylated covalently-linked homodimer of 64 kilodaltons (kDa). NOG binds with various degrees of affinity to BMP-2, -4, -5, -6 and -7, GDF-5, -6 and vegetally localized protein (Vg)-1, but not other members of the TGF-beta family of peptides. Crystallographic studies of NOG and BMP-7 reveal that noggin can inhibit BMP-7 signaling by blocking the molecular interfaces of the binding epitopes for both type I and type II BMP receptor, thus making it a very efficient antagonist. The embryonic lethality of NOG-null mice has not allowed the definition of its role in adult tissues. As such, NOG has been investigated in the context of bone formation and developmental processes. Importantly, homozygous null mutations of the NOG gene result in serious developmental abnormalities, such as failure of the neural tube formation, dysmorphic development of the axial skeleton and joint lesions. The importance of the NOG gene was also confirmed by human studies demonstrating that different heterozygous missense mutations of
NOG coding sequence may lead to proximal symphalangism and multiple synostosis syndrome\textsuperscript{284, 285}. The investigation of NOG in cancer progression is only recently beginning to emerge. For instance, a causative link with dedifferentiation phenomena has been established in certain contexts, such as in the case of the BMP6-NOG axis in advanced prostate cancer\textsuperscript{286}, as well as with the development of osteolytic bone metastases that characterize the dissemination pattern of most prostate cancers\textsuperscript{287, 288}.

Chordin is secreted as a glycosylated homodimer of 120 kDa, is characterized by four CR domains, which are the sites of interaction with BMPs, and binds specifically to BMP-2, -4 and -7\textsuperscript{289}. Chordin is regulated through interactions with other ECM proteins. The chordin-BMP complex is a substrate for the zinc metalloprotease BMP1/tolloid, which cleaves chordin and liberates BMPs to the extracellular space\textsuperscript{290}. The proteolytic activity of BMP1/tolloid is specific for chordin and does not modify the action of other BMP antagonists, such as NOG or members of the CAN family\textsuperscript{291}. Chordin-null mutations result in stillborn mice, with pharyngeal and cardiovascular deficits. Double-mutant noggin/chordin mice display defects in the development of the forebrain, eye and facial structures, and exhibit disrupted mesoderm development and left to right patterning, further indicating that they are both required for the proper establishment of the three body axes in the mouse embryo\textsuperscript{292}. Chordin-like CR domains are present in a growing number of extracellular proteins, including Chordin like-1, Chordin like-2, Procollagen type II, Kielin, Amnioless, Nell, Crossveinless-2, Crim-1, and members of the CCN family. Although, these chordin-like CR proteins may potentially act as BMP antagonists, some may have BMP-promoting effects (i.e. Crossveinless-2), or possess a wide variety of protein domains, which may signal more important biological functions, independent of BMP antagonism (i.e. CCN proteins)\textsuperscript{293}. The role of chordin -if indeed any- in human cancer is particularly underexplored. It is believed that it may be specifically involved with reduced motility of cancer cells, and significant underexpression has been noted in certain ovarian tumors\textsuperscript{294}.

Twisted gastrulation (TSG), a secreted glycoprotein with a molecular weight of 23.5 kDa, was originally identified in \textit{Drosophila}. The gene encodes a secreted protein with two evolutionary conserved CR domains, of which the amino-terminal one is necessary for direct interaction between TSG and BMP-2 and -4, whereas the carboxy-terminal one directly interacts with chordin\textsuperscript{295}. Phenomenally, TSG can display both BMP antagonist and agonist functions\textsuperscript{276}. As a BMP antagonist, TSG binds directly with BMP-2 or -4 or with a BMP-chordin preformed complex, generating a trimolecular structure, which is more efficient than its individual components in inhibiting BMP signaling\textsuperscript{295-297}. As a BMP agonist, TSG can enhance chordin cleavage by BMP1/tolloid\textsuperscript{298, 299}. TSG is involved in many developmental processes, including maturation of bone marrow cells and proper skeletal development\textsuperscript{240, 276}, and a causative link with cancer progression, if any, is yet to be established\textsuperscript{300}. 
CAN is a family of secreted glycoproteins capable of covalently binding to and suppressing various BMPs. At least nine members of the CAN family have been described: gremlin, sclerostin, differential screening-selected gene aberrative in neuroblastoma (DAN), uterine sensitization associated gene (USAG-1), cerberus, caronte, coco, protein related to dan and cerberus (PRDC) and dante. For a more detailed description in each one of these, refer to the comprehensive review by Gazzerro and Canalis (2006). Protein sequence alignments reveal that these proteins share a common carboxy-terminal CR domain consisting of nine cysteines, outside of which, they display very little homology. Here, we will briefly refer to only two of these BMP antagonists, gremlin and sclerostin, which have been explored in more detail compared to the remaining ones. The gremlin gene was cloned from a Xenopus ovarian library for its axial patterning activities and encodes for a 28 kDa glycosylated peptide, which can bind to BMP-2, -4, and -7 with high affinity. During embryonic development, grelin is expressed in the limb bud mesenchyme, lung, and kidney. As such, homozygous null deletion of the gremlin gene in mice results in serious developmental limb abnormalities, lung airway defects, kidney agenesis and lethality.

Interestingly, the gremlin/BMP7 axis has been extensively related to the pathophysiology of various non-neoplastic fibrotic diseases, such as pulmonary fibrosis, diabetic nephropathy, and pauci-immune glomerulonephritis, as well as neoplastic fibrotic/desmoplastic ones (i.e. squamous cell carcinoma). Encoded by the SOST gene, Sclerostin is a secreted glycoprotein that exhibits very low homology with other members of the CAN family, and may bind to BMP-2, -4, -5, -6, -7 with weak affinity. SOST gene mutants are responsible for sclerosteosis and Van Buchem disease, two congenital dysplasias characterized by progressive bone thickening and generalized osteosclerosis. Sclerostin, can also directly bind to Wnt co-receptors and inhibit the Wnt/beta-catenin canonical signaling pathway, which actually may be the most dominant mechanism of its implication in tumorigenesis.

1.3.4. Sequestration of BMPs in the Extracellular Matrix

BMPs are not diffused in the ECM without encountering biological barriers. Following their secretion, various ECM components may participate in protein-protein interactions with BMPs and either sequester or facilitate binding to their receptors and subsequent downstream signaling. For instance, it has been shown that the ECM component Matrix Gla Protein (MGP) may normally bind to and inhibit BMP signals upstream of the corresponding receptors, and quite unexpectedly an MGP-knockout mouse model presented excessive BMP-mediated angiogenic sprouting. Given the above, it is plausible to generally presume that extracellular proteolytic cascades could potentially alter the degradome of the exoproteome, as well as the molecular dynamics between ECM and BMP growth factors that are released in the extracellular microenvironment. A prominent example and well-studied ECM component, known
to be interacting with BMPs and TGF-beta is decorin, an archetypal member of the small leucine-rich proteoglycan gene family. Decorin has a very broad interactome, which not only encompasses matrix structural components (such as collagens and growth factors), but also a wide variety of proteases, which can readily process decorin, consequently altering its structure and potential of interacting with other factors. For example, extracellular proteolytic systems like high-temperature requirement-1 (HTRA1) and -3 (HTRA3) are responsible for cleaving specific peptide compartments of both decorin and MGP, which may, in turn, alter their structure in a manner that sequesters TGF-beta and BMPs in the ECM. As such, HTRA1 and HTRA3 may indirectly serve as BMP suppressors/inhibitors, despite the fact they are not specific BMP antagonists. As expected, interactions of BMPs with ECM may add another level of complexity in the extracellular regulation of the BMP pathway, besides the one described with the traditional BMP antagonism.

1.3.5. Implications of BMP Pathway in Development and Healthy Tissue Homeostasis

The literature on the implication of the BMP pathway in developmental process and healthy tissue homeostasis is vast, and a detailed description is beyond the scope of the current thesis. BMPs were initially identified by their capacity to induce endochondral bone formation, and their role in skeletogenesis remains the most well-characterized part of their regulatory impact in development. In the limb bud, they interact with FGF4 and Sonic Hedgehog, reducing limb bud expansion and inducing the formation of chondrocyte and osteoblast precursors. After birth, BMPs keep playing important roles in the maintenance of bone mass. They induce the differentiation of bone-marrow stromal cells toward the osteoblastic lineage, therefore increasing the pool of mature bone-forming cells, and enhance the differentiated function of osteoblasts. The relevance of BMP effects in bone tissue has been confirmed through mouse models, in which distinct BMP ligands and their downstream effectors were genetically modified in a tissue-specific fashion. A pivotal role of BMPs in fracture healing has also been proposed.

The establishment of the dorsal–ventral (DV) axis is one of the first key events during early development. It has been shown that secreted ligands from the dorsal blastopore lip (Spemann organizer) are essential for proper embryonic patterning, and further genetic studies collectively suggested that gradients of BMPs and their antagonists may drive differentiation of ectodermal cells into central nervous system, neural crest, and the epidermis. Besides their implicit roles in early patterning of the central nervous system (CNS), the determination of neural cell fates and regulation of oligodendrocyte maturation
during CNS development seem to also be mediated through BMPs 333, 334. After birth, regenerating roles during CNS injury are proposed to be regulated through BMPs, among other pathways 333.

Interestingly, various constituents of the BMP signaling pathway have been found to be expressed in an autocrine or paracrine aspect in a wide plethora of postnatal adult tissues, in close conjunction with members of other pathways, especially the Wnt pathway 277. Besides the observation that certain BMP antagonists may also directly target the Wnt pathway 311, it is now accepted that the BMP pathway itself, is a potent negative regulator/suppressor of the Wnt pathway, and these reciprocal interactions may be responsible to maintain differentiation equilibriums in cells positioned in a variety of contexts, such as the maintenance of the stem cell niche in the bottom of intestinal crypts 335-337, the interactions of epithelial components of the kidney with the mesenchyme 338, 339, the specification of intrapulmonary stem cells 340, and the dynamic microenvironment around the hair follicle that regulates hair growth 254, 341.

1.3.6. Implications of BMP Pathway in Cancer Development and Progression

Cancer very often involves deregulation of developmental pathways, including the BMP signaling pathway, and such involves disturbances from the level of ligands/antagonists down to the level of receptors and the entire downstream repertoire 342. However, the full picture on the causative role of BMP pathway in cancer remains rather elusive. On one hand, there are well-established tumor-promoting effects linking BMPs, receptors and interacting molecules to carcinogenesis and tumor progression, but BMPs can sometimes actively play the role of tumor suppressors 342, 343. This can be easily explained from the profound complexity of the organization of the BMP pathway, as described in the previous sections, and a widely- and reasonably-accepted consensus supports the notion that the effects of the BMP pathway in cancer development and progression are considered as context-dependent 344, 345. In concordance with an earlier viewpoint that TGF-beta signaling has both tumor-promoting and tumor-suppressive effects 346, it seems that the BMP pathway itself follows the exact same pattern, further designating its role as another molecular “Jekyll and Hyde” of tumorigenesis 344.

Here, we will summarize a few important points that denote causative implication of various BMP components in the regulation of colorectal cancer (CRC) 345, a type of cancer that is used as a model system in this thesis. Studies on families with inherited polyposis syndromes that eventually develop various forms of CRC have collectively accused several signaling pathways as tumor-initiating ones, with most prominent example that of the Wnt pathway (i.e. APC mutations) 345. A further signal transduction pathway that has been identified as an important barrier to intestinal tumor initiation is the BMP pathway.
that may be related with the pathogenesis of many histological/molecular pathways observed in CRC\textsuperscript{345}. A detailed analysis on these molecular pathways is presented in chapter 1.4.3.

The BMP pathway is implicated in the initiation of CRC through inactivating mutations of BMPR1A in patients with juvenile polyposis (JP)\textsuperscript{347}. Mutations in SMAD4 are also identified in such patients\textsuperscript{348}, and along with those at the receptor level, they account for approximately half of all cases of JP\textsuperscript{349}. JP is an autosomal dominant hamartomatous polyposis syndrome with an increased risk for the development of CRC\textsuperscript{350}. The distinctive characteristic of hamartomas compared with adenomas is their abnormal mesenchymal component. In particular, the hamartomas in JP have an opulent mesenchymal stroma with inflammatory infiltrates of lymphocytes and plasma cells\textsuperscript{351}. Interestingly, in-depth studies on the SMAD4 germ-line mutation in the case of JP have demonstrated that the loss was observed in both cancer cells and CAFs, but not in the inflammatory infiltrates\textsuperscript{352}. In concordance with the previously described “memory-from-progenitor” hypothesis regarding the distinct CAF subsets in the tumor microenvironment (refer to chapter 1.1.2.4.), this may strongly suggest that the mutant cancer cells and CAFs could be derived from a single mutant epithelial clone as a result of EMT for the former, since it seems particularly impossible that an independent hit of SMAD4 co-occurred in a stromal cell clone.

Recent developments in CRC genetics have led to further understanding on the implications of BMP pathway in sporadic CRC\textsuperscript{353}. In the sporadic form, BMP is also suppressed\textsuperscript{345}, and multiple levels have been proposed in this context. At the extracellular level, BMP2, BMP3, BMP4 and BMP7 have been found to be growth-suppressive\textsuperscript{354-356}. However, the expression of BMP4 and BMP7 is paradoxically increased with disease progression and correlates with a poor prognosis\textsuperscript{357, 358}. Also, it has been demonstrated that the paracrine secretion of BMP antagonists, such as GREM1 may impair BMP2/7 signaling and promote tumor cell motility and proliferation\textsuperscript{309}. At the receptor and signal transduction level, there was no evidence for loss of BMPR1A, as opposed to JP-derived CRC\textsuperscript{355}, but some recent data suggested genomic mutations of SMAD4 and BMPR2\textsuperscript{359, 360}. SMAD4 is very frequently deleted in sporadic CRC but the biological significance of this has always been attributed to loss of TGF-beta signaling rather than BMP signaling\textsuperscript{361}. It was shown that the loss of phosphorylated SMAD1/5/8 mostly occurs during the adenoma-cacinoma sequence\textsuperscript{360}. It is now strongly believed that BMPR2 inactivating mutations mostly occur through the microsatellite instability pathway\textsuperscript{360}, while CRCs developed without defects in mismatch repair mechanism mostly depend on developing alternative strategies for BMP suppression, principally regulated at the extracellular and the epigenetic level\textsuperscript{345, 362}. 
1.4. COLORECTAL CANCER

1.4.1. General Information

Colorectal cancer (CRC), also known as large bowel cancer, is cancer arising in the colon and/or rectum, two distinct anatomical localizations which otherwise are considered as the same type of cancer, due to sharing identical genetic characteristics \(^{353}\). Worldwide, CRC is the second most common cause of cancer-related death with an annual incidence of 1.2 million and an annual mortality over 600,000 \(^{363,364}\). Clinical signs of CRC depend on tumor localization, and whether metastasis has already occurred. Common symptoms are not informative and only indicative of CRC, and include constipation, blood in the stool, anorexia, abdominal pain, loss of weight, and possibly nausea and vomiting \(^{365,366}\). The treatment for CRC is distinguished to curative and palliative therapy, with the decision usually depending on the patient's health and preferences, and the stage of the tumor. When colorectal cancer is caught early, surgical treatment, involving removal of a piece of intestine away from the tumor margins, may completely cure the disease. However, the disease is very rarely caught at this stage, especially due to the non-specific nature of the symptoms that do not alert the patients, and as such, metastases are most likely present, and palliative treatment is performed \(^{367}\). Besides certain well-characterized, possibly clinically silent, underlying factors, which are known to predispose to CRC (i.e. genetic syndromes caused by specific mutations, inflammatory bowel disease etc.) \(^{368-370}\), a considerable number of lifestyle habits have been found responsible for the development of CRC, including smoking, obesity, lack of physical exercise and alcohol consumption \(^{371}\). As such, preventive cancer medicine aims to eliminate the risk derived from these factors not only through surveillance, but also with invasive lifestyle changes \(^{372}\).

1.4.2. Biomarkers in Colorectal Cancer

Due to the slow and multi-stage progression of CRC, as well as the general absence of symptoms at an early stage, it is estimated that around 40% of patients have overt metastases at presentation \(^{373}\). This observation provides the rationale for the most important clinical need regarding CRC management, which is the identification of non-invasive biomarkers for early detection of the disease. This additionally includes the detection of pre-malignant and clinically significant polyps and adenomas, which are considered critical for reducing both incidence and mortality. When diagnosed early, the 5-year survival rate for CRC is 90-95%, indicating a high curative rate. In comparison, when CRC is detected at later stages, the 5-year survival rate is significantly less (5-10%) \(^{374}\).
Currently, the fecal occult blood test (FOBT), fecal immunochemical test (FIT), colonoscopy and sigmoidoscopy are the only clinically accepted diagnostic tests of CRC. The FOBT and FIT are used to detect the presence of heme or blood in stool and whilst these test have relatively low cost, they are regarded as having poor sensitivity for early stage disease. Because the presence of blood in stool is also not specific for CRC, the FOBT and FIT suffer from relatively high false positive rates. In contrast, while colonoscopies compensate for these disadvantages (i.e. they have high specificity for the disease), the procedure is highly invasive and expensive. Two of the most widely-known serum protein biomarkers for gastrointestinal malignancies, including CRC, are the carbohydrate antigen CA19-9 and carcinoembryonic antigen (CEA). CEA, while useful for monitoring recurrence of CRC, exhibits a specificity for cancer of 87%, but its sensitivity (35%) is too low to be useful for detection of CRC in an asymptomatic screening population. Similarly, CA19-9 has limited utility as a diagnostic marker due to its lack of specificity for malignant disease.

Currently, but still in research phase, are several promising DNA diagnostic biomarkers for CRC. One of the most promising ones, methylated septin 9 (mSEPT9) measured in plasma may provide high detection rates of CRC. In a case-control study of 92 controls and 92 CRC patients, mSEPT9 was reported to have higher sensitivity and specificity than either FOBT or CEA, and such phenomenal performance was further validated in subsequent clinical trials, despite that fact that detection rate for adenomas was relatively low. A stool-based DNA test consisting of a panel of four methylated genes (BMP3, NDRG4, VMN, TFPI2) and a mutant form of RAS has also been developed. In comparison to mSEPT9, the stool-based DNA test performed, in general, similarly well, however it was markedly more sensitive in detecting early stage disease (i.e. adenomas). In a large clinical trial conducted in an asymptomatic screening population (n=7,941), the stool-based DNA test was significantly more sensitive for diagnosing CRC than plasma mSEPT9 (sensitivity of 87% versus 60% for CRC overall).

Something that remains a major challenge even today though, is the identification of serological markers for CRC diagnosis, especially those that could be used as an easily deployed screening tool and could therefore substitute the current expensive and displeasing (for the patients) methods. Despite being particularly early and no promising markers have yet risen, a couple of proteomic studies have been utilizing various *in vitro* and *in vivo* platforms to address this issue. A conclusive consensus that could be drawn from all these studies is that through the constant improvement of method development along with technological advancements that seem to be progressing very rapidly in the modern *omics* era, we will probably witness the emergence of highly informative CRC markers or panels of biomarkers in the near future.
In conclusion, the field of CRC biomarker discovery is extensively active and a vast plethora of novel biomarkers is proposed to address the current clinical needs. However, the lack of success of most follow-up studies and lack of consistency between these studies remains an elusive matter of debate. Possible factors participating in such discrepancies may be small cohort sizes, cohort composition, assay reproducibility, biomarker stability, differences in sample handling and/or processing procedures, and over-representation of late stage disease patients which can bias biomarker sensitivity estimates.

1.4.3. Genetic Pathways and Molecular Subtypes of Colorectal Cancer

While CRC is mostly identified in the sporadic form, it is now known that it may also occur in the context of inflammatory bowel disease (IBD), such as ulcerative colitis (UC), or a wide variety of genetic syndromes, such as FAP, JP, and Lynch syndrome [that may cause hereditary non-polyposis colon cancer (HNPCC)]. All the latter usually serve as predisposing factors by preparing the context leading to CRC development through specific molecular/genetic pathways, which also reflect to unique histological features in the developing tumors. However, they together comprise a very limited portion of all CRC cases, and as such, are not the focus of the current thesis.

The development of sporadic CRC is caused by the accumulation of genetic and epigenetic changes, but the description of distinct molecular subtypes has proven a particularly difficult task in the case of CRC, due to the phenomenal heterogeneity of the disease. However, there have been certain molecular drivers that have provided some insights in understanding disease progression and these have been generally used as organizing principles for creating molecular subtypes in CRC.

A prominent molecular driver in sporadic CRC is the status of the DNA mismatch repair machinery, a system responsible for recognizing and repairing erroneous insertions, deletions or defective incorporations of bases, during DNA replication/recombination. Approximately 80% of CRC patients undergo a well-characterized series of molecular events, collectively described as adenoma-carcinoma sequence, involving chromosomal aberrations and mutations in several genes, such as APC, KRAS, P53 and DCC with no defects in DNA mismatch repair. This pathway is also referred to as chromosomal instability (CIN) pathway and all associated aberrations (i.e. mutations) are attributed to it. However, the remaining 20% of CRC patients undergoes a secondary molecular pathway, which causes genetic instability in microsatellite loci (MSI+) attributable to alterations in the DNA mismatch repair genes, such as MLH1, MSH2, MSH6 and PMS2. These latter cancers are considered as replication error-deficient (RER+), while the CIN+/MSI- as replication error-proficient (RER-) It is evident that these two molecular pathways have different and
independent mechanisms of tumor development and progression. For instance, RER+ cancers have microsatellite sequence alterations in IGFR2 and TGFR2\(^2\), and frequently acquire the first hallmark of cancer, i.e. sustaining of proliferation signaling, through BRAF mutations rather than KRAS, as happens in the case of RER- cancers\(^2\). Also, RER+ cancers have better prognosis, which may be attributed to increased apoptosis, due to extensive DNA damage and instability sensing within cells\(^2\). As such, the mechanism by which the third hallmark of cancer, i.e. evasion from cell death, is acquired is again quite distinct between those two. In brief, the P53 gene is targeted in the case of RER-, but apoptosis regulator gene (BAX), a member of the B-cell lymphoma-2 (Bcl-2) gene family that promotes apoptosis by heterodimerizing with and antagonizing the Bcl-2 protein\(^2\), is targeted in the case of RER- cancers\(^2\).

Another prominent molecular driver is the epigenetic disturbances in the course of CRC development\(^4\). Approximately 15 years ago, an independent molecular pathway for the development and progression of CRC, involving aberrant hypermethylation of multiple gene promoters, known as CpG island methylator phenotype (CIMP+), was introduced\(^2\). Ever since, a vast plethora of studies has concluded that certain CRCs develop through a CIMP-low pathway, while others involve the epigenetic suppression of multiple tumor-suppressor genes through DNA methylation\(^4\). CIMP+ do not involve APC gene inactivation as an early event during tumorigenesis\(^2\). Histologically, CIMP+ patients tend to develop a phenotype known as serrated adenoma/adenocarcinoma, which is prominently characterized by branching and/or aberrant dilation of the neoplastic crypts\(^2\), and may further progress through the inactivation of critical genes\(^2\). Interestingly, CIMP+ and MSI pathways do not seem to act independently, as it has been shown that the CIMP+ pathway may downregulate mismatch repair genes, for instance through hypermethylation of the hMLH1 promoter. CIMP+ RER+ cancers tend to very quickly achieve BRAF and KRAS mutations and usually present with mucinous histology\(^4\).

The most efficient and broadly accepted method for identifying molecular subtypes in cancer is currently the hierarchical clustering of gene expression profiles in large patient cohorts\(^4\). Such approaches have also been performed in the case of CRC and have proposed numerous classifiers, enough of which identify the above-discussed microsatellite stability status and the hypermethylator phenotype\(^4\). However, none of these has yet established a robust consensus for molecular subtyping of CRC, despite that results (in some of those) are quite promising given they were validated in independent patient cohorts\(^4\). Nevertheless, efforts towards better understanding of the molecular pathways underlying this hererogenous disease should definitely continue, especially by learning from the very recent example of breast cancer, where collective efforts resulted in the development of one clinically-helpful molecular classification system\(^3\).
1.4.4. Colon Cancer Cell Lines: Can They Actually Recapitulate the Disease?

Although certain differences exist, it is generally accepted that genomic and transcriptional profiles of cancer cell lines may recapitulate the most leading features of primary tumors\textsuperscript{182-184}. The most significant argument against this notion is the lack of tumor microenvironment, which forces cancer cells grow in a context, biologically irrelevant to the one observed in actual tumor tissues. As such, cancer cells tend to get (epi)genomically altered, after certain passages\textsuperscript{178}.

Colon cancer is a highly heterogenous disease, and this feature apparently renders the utilization of colon cancer cell lines for \textit{in vitro} investigations a rather tricky process. It follows that results obtained from studies focusing on certain signaling pathways would be very dependent on the selection of the proper cell lines, and their conclusions would probably reflect to specific subgroups of CRC patients that share a similar molecular portrait with the cell lines worked with. Over the years, a considerable number of commercially available CRC cell lines, with distinct genotypic and phenotypic features, as well as biological behaviors have been generated. However, only recently, there have been attempts to globally characterize their molecular features.

In one such study\textsuperscript{436}, a total of 24 commonly used cell lines were investigated for the three commonly encountered, partly overlapping, molecular phenotypes reflecting different forms of DNA instability, which are chromosomal instability, microsatellite instability and CpG island methylator phenotype. Interestingly, cell lines included in that study\textsuperscript{436} varied in appearance (a few of them were “dense island”-forming, i.e. EB, FRI, IS3, LS1034, SW1116, and V9P, with the remaining mostly forming monolayers), and growth characteristics (a few of them were fast-growing, i.e. Caco-2, Colo-320, DLD-1, HCT-15, HCT-116, HT-29 and TC71, with the remaining having doubling times >24h). This phenotypic diversity could be attributed, in part, to the diverse repertoire of CRC molecular pathways, which were subtyped according to CIN, MSI and CIMP profiles, as well as the mutation status of five critical genes (KRAS, BRAF, PIK3CA, TP53 and PTEN)\textsuperscript{436}, using comprehensive and sensitive panels of discriminatory biomarkers\textsuperscript{437-443}. Interestingly, a remarkable heterogeneity was also observed at the protein level, from our group\textsuperscript{444}, during proteomic analysis of various secretomes. A variable number of secreted proteins have been observed with only mild overlap between the different CRC cell lines, and these differences could most probably reflect the phenomenal diversity and heterogeneity described above. All these data are summarized on Table 1.1.

Overall, it is very important that scientific discoveries in \textit{in vitro} settings should take into account the genotypic and phenotypic plasticity of CRC, and conclusions should always be made on the basis of specific cell line selection and experimentation.
Table 1.1. Molecular portrait of 24 commercially available, commonly used CRC cell lines [adapted and modified from Ahmed et al. (2014) 436]

<table>
<thead>
<tr>
<th>Cell line</th>
<th>MSI status</th>
<th>CIN status</th>
<th>CIMP status</th>
<th>KRAS</th>
<th>BRAF</th>
<th>PIK3CA</th>
<th>PTEN</th>
<th>TP53</th>
<th># of Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO-115</td>
<td>RER+</td>
<td>CIN-</td>
<td>CIMP+/+</td>
<td>WT</td>
<td>M</td>
<td>WT</td>
<td>M</td>
<td>WT</td>
<td>ND</td>
</tr>
<tr>
<td>DLD-1</td>
<td>RER+</td>
<td>CIN-</td>
<td>CIMP+/+</td>
<td>M</td>
<td>WT</td>
<td>M</td>
<td>WT</td>
<td>M</td>
<td>551</td>
</tr>
<tr>
<td>HCT-116</td>
<td>RER+</td>
<td>CIN-</td>
<td>CIMP+/+</td>
<td>M</td>
<td>WT</td>
<td>M</td>
<td>WT</td>
<td>WT</td>
<td>1326</td>
</tr>
<tr>
<td>HCT-15</td>
<td>RER+</td>
<td>CIN-</td>
<td>CIMP+/+</td>
<td>M</td>
<td>WT</td>
<td>M</td>
<td>WT</td>
<td>M</td>
<td>ND</td>
</tr>
<tr>
<td>LoVo</td>
<td>RER+</td>
<td>CIN-</td>
<td>CIMP/-</td>
<td>M</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>1085</td>
</tr>
<tr>
<td>LS174T</td>
<td>RER+</td>
<td>CIN-</td>
<td>CIMP/-</td>
<td>M</td>
<td>WT</td>
<td>M</td>
<td>WT</td>
<td>WT</td>
<td>954</td>
</tr>
<tr>
<td>RKO</td>
<td>RER+</td>
<td>CIN-</td>
<td>CIMP+/+</td>
<td>WT</td>
<td>M</td>
<td>M</td>
<td>WT</td>
<td>WT</td>
<td>1269</td>
</tr>
<tr>
<td>SW48</td>
<td>RER+</td>
<td>CIN-</td>
<td>CIMP+/+</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>ND</td>
</tr>
<tr>
<td>TC71</td>
<td>RER+</td>
<td>CIN-</td>
<td>CIMP/-</td>
<td>M</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
<td>M</td>
<td>ND</td>
</tr>
<tr>
<td>Caco-2</td>
<td>RER-</td>
<td>CIN+</td>
<td>CIMP-/</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
<td>ND</td>
</tr>
<tr>
<td>COLO-320</td>
<td>RER-</td>
<td>CIN+</td>
<td>CIMP-/</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
<td>1868</td>
</tr>
<tr>
<td>EB</td>
<td>RER-</td>
<td>CIN+</td>
<td>CIMP/-</td>
<td>M</td>
<td>WT</td>
<td>M</td>
<td>WT</td>
<td>WT</td>
<td>ND</td>
</tr>
<tr>
<td>FRI</td>
<td>RER-</td>
<td>CIN+</td>
<td>CIMP/-</td>
<td>M</td>
<td>WT</td>
<td>M</td>
<td>WT</td>
<td>M</td>
<td>ND</td>
</tr>
<tr>
<td>HT-29</td>
<td>RER-</td>
<td>CIN+</td>
<td>CIMP+/+</td>
<td>WT</td>
<td>M</td>
<td>M</td>
<td>WT</td>
<td>M</td>
<td>ND</td>
</tr>
<tr>
<td>IS1</td>
<td>RER-</td>
<td>CIN+</td>
<td>CIMP/-</td>
<td>M</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
<td>ND</td>
</tr>
<tr>
<td>IS3</td>
<td>RER-</td>
<td>CIN+</td>
<td>CIMP/-</td>
<td>M</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
<td>ND</td>
</tr>
<tr>
<td>LS1034</td>
<td>RER-</td>
<td>CIN+</td>
<td>CIMP/-</td>
<td>M</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
<td>ND</td>
</tr>
<tr>
<td>NCI-H508</td>
<td>RER-</td>
<td>CIN+</td>
<td>CIMP/-</td>
<td>WT</td>
<td>M</td>
<td>M</td>
<td>WT</td>
<td>M</td>
<td>ND</td>
</tr>
<tr>
<td>SW1116</td>
<td>RER-</td>
<td>CIN+</td>
<td>CIMP/-</td>
<td>M</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
<td>973</td>
</tr>
<tr>
<td>SW480</td>
<td>RER-</td>
<td>CIN+</td>
<td>CIMP/-</td>
<td>M</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
<td>1549</td>
</tr>
<tr>
<td>SW620</td>
<td>RER-</td>
<td>CIN+</td>
<td>CIMP/-</td>
<td>M</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
<td>1376</td>
</tr>
<tr>
<td>SW948</td>
<td>RER-</td>
<td>CIN+</td>
<td>CIMP/-</td>
<td>M</td>
<td>WT</td>
<td>M</td>
<td>WT</td>
<td>M</td>
<td>ND</td>
</tr>
<tr>
<td>V9P</td>
<td>RER-</td>
<td>CIN+</td>
<td>CIMP-/</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>M</td>
<td>ND</td>
</tr>
<tr>
<td>WiDr</td>
<td>RER-</td>
<td>CIN+</td>
<td>CIMP+/+</td>
<td>WT</td>
<td>M</td>
<td>M</td>
<td>WT</td>
<td>M</td>
<td>961</td>
</tr>
</tbody>
</table>

Non-standard abbreviations: MSI, microsatellite instability; RER, replication error; CIN, chromosomal instability; CIMP, CpG island methylator phenotype; M, mutated; WT, wild type; ND, not-determined
1.5. STUDY AIMS, RATIONALE, AND HYPOTHESIS

1.5.1. Rationale

Cancer cells of solid carcinomas migrate in collective configuration, immediately after breaching the basement membrane and escaping their in situ state. Accumulating evidence supports that the penetration into the underlying normal tissues and the migration of the cancer cell compartment towards lymphatic or blood vessels is achieved from the downward invasive margins, also known as the cancer invasion front. It has been long observed that cancer cells that have adapted in the periphery of such invasion fronts may obtain stem-like properties, involving the loss of their epithelial shape/polarity with a parallel gain of very motile activity. Histological observations have also documented that these mesenchymal and aggressive characteristics are correlated with prominent dedifferentiation, especially when these peripheral cancer cells are compared with cancer cells positioned in the tumor core. Such dedifferentiated phenotype has been shown to be achieved through the deployment of sophisticated and complicated cell biological programs, such as the epithelial-to-mesenchymal transition (EMT). EMT is a cell-biological program that is often encountered in various developmental processes, and is shown to permit the developmental plasticity and coordination that are both demanded for proper tissue/organ formation. It, therefore, appears that cancer cells of the invasion front are capable of reprogramming their gene- and protein-expression machinery, to exert an EMT-like phenotype, and further obtain the aforementioned stem-like properties, which facilitate cell dissemination and metastatic propagation. Among the most direct consequences of EMT exploitation, selectively and almost exclusively by cancer cells lining invasion front, is the so-called local invasion (i.e. invasion into and migration through the adjacent stromal tissues), and in an exaggerated paradigm, the complete detachment from the tumor core and the formation of isolated nest-like aggregates, also known as tumor budding.

A very prominent question, which would be incremental to illuminate the unknown mechanisms of the metastatic cascade, is how is this dedifferentiation reprogramming (EMT) achieved at the molecular level, in this spatially-organized area of the tumor? Collective evidence from the investigation of the EMT program argues that the major determinant for EMT initiation in developing tissues and organs is often an extensive set of extracellular, contextual signals originating from the cell microenvironment. Accordingly in the neoplastic tissue, EMT initiation at the level of the invasion front is believed to also occur through context-dependent mechanisms, originating from the tumor microenvironment. These contextual signals could be released from recruited cancer-associated stromal cells constituting the tumor microenvironment, and subsequently act in a paracrine, juxtacrine or even endocrine manner to induce EMT on selective cancer cells lining the invasion front. However, the current dogma supports a more complicated model,
due to the documented heterogeneity that characterizes the tumor microenvironment. This model does not actually deviate a lot from what has already been described, but it rather states that there is a plethora of contextual signals that could act as suppressors or promoters of EMT in the tumor-host cell interface, at any given time-point. As such, the definite driver of EMT along with local invasion and tumor budding formation is actually a switch-type balance, shifted from accumulation of EMTsuppressing (or MET-promoting) to EMT-promoting contextual signals at the cancer invasion front. Unfortunately, the precise nature and origin of these contextual signals has been rather unexplored, with only a few exceptions, such as numerous indications for the involvement of signal transduction pathways, such as TGF-beta and Wnt. It follows that comprehensive delineation and deciphering of these contextual signals would not only assist in our understanding on these molecular and cellular processes, but it would potentially support the development of more targeted therapies.

As already stated, the source of such EMT-regulating contextual signals is the tumor microenvironment, which is a highly-complex and heterogeneous tissue. Various lines of evidence demonstrate that the tumor-associated stromal cells supply and support the developing tumors by recruitment and proliferation of pre-existing stromal, progenitor and stem cells, either these are derived from the adjacent, quiescent, healthy tissue, or the bone-marrow. The most abundant cell types that constitute the context at sites of the cancer invasion front are inflammatory/immune infiltrates and cancer-associated fibroblasts (CAFs). CAFs are, in fact, reprogrammed derivatives resembling the so-called myofibroblasts, also known as activated fibroblasts seen at sites of wound healing. CAFs may rise as a quite distinct subpopulation of the tumor stroma from a wide variety of cell-sources, mostly involving normal quiescent fibroblasts, mesenchymal stem cells, and other bone marrow progenitors. Histological observations in advanced tumor margins, especially around sites of EMT induction and tumor budding formation, have regularly reported the dominant presence of desmoplastic reaction (i.e. CAF aggregates with prominent extracellular matrix deposition), as a proponent feature of this microenvironment. Concrete and accumulating evidence has conceptualized that this desmoplasic interface might serve as a prominent source of tumor promoting mediators, which may also include EMT-regulatory elements.

1.5.2. Hypothesis

Taking together three major constituents of this rationale, in particular: (a) metastatic dissemination from the invasive margins of tumors is mediated through dedifferentiation reprogramming (i.e. EMT), (b) EMT is primarily dictated by contextual signals, originating from the tumor microenvironment, and that (c) this microenvironment is mostly characterized by intense desmoplastic reaction (i.e. multiple CAF aggregates
and aberrant ECM deposition) in many solid carcinomas, we have formulated the hypothesis that reciprocal paracrine interactions between cancer cells and cancer-associated fibroblasts may result in a contextual, secreted, molecular signature, which may orchestrate local invasion and tumor budding formation in the proximal invasion front.

1.5.3. Objectives and Study Pipeline

To gain important insights into this complex hypothesis (Fig. 1.10A.), we propose a 5-step experimental pipeline (Fig. 1.10B.). We begin with the comprehensive profiling of cancer cell-CAF interactions, to gain a global perspective of the exoproteome present in the desmoplastic interfaces. We then narrow this secreted signature down to a smaller number of promising candidates, some of which are selected for further analysis and investigation.

The specific objectives are summarized as follows:

1. To generate a molecular signature/profile of the desmoplastic invasion front, by performing proteomic analysis in cocultures of cancer cells with cancer-associated fibroblasts.

2. To identify molecular themes involved in local invasion and tumor budding formation, by performing integrative pathway analysis in the proteomic signature obtained in objective #1.

3. To validate molecular drivers of interest (i.e. GREM1, FST and HTRA3), proposed in objective #2, by demonstrating their overrepresentation in various in vitro and tissue-wide platforms.

4. To characterize the protein expression patterns of the molecular drivers of interest (i.e. GREM1, FST and HTRA3) in cancer patients, and to additionally provide phenotypic correlation with presence of tumor budding in desmoplastic invasion fronts, through immunohistochemistry.

5. To provide: (a) contextual/phenotypic association between GREM1-expressing desmoplastic stromata and EMT-undergoing cancer cells, through immunohistochemistry, and (b) mechanistic association between GREM1 function and EMT in cancer cells, through in vitro functional assays.
FIGURE 1.10. (A) Hypothesis scheme and (B) study design. EMT, epithelial-to-mesenchymal transition; MET, mesenchymal-to-epithelial transition.
CHAPTER 2
Delineating the Secretome of the Colorectal Cancer Desmoplastic Invasion Front

Sections of this chapter have been published in Oncotarget and Molecular Cancer Research:


Copyright permission has been granted.
2.1. INTRODUCTION

Emerging evidence suggests that local invasion and tumor budding formation at the level of the invasion front is a context-dependent process, regulated from the paracrine signals secreted from the proximal cancer-associated stroma\textsuperscript{145}. In the case of colorectal cancer, a type of cancer where tumor-budding formation and tumor invasion has been a subject of great scientific interest\textsuperscript{145, 154}, it has been shown that such peritumoral stroma is desmoplastic in nature, and consists of the so-called “cancer-associated fibroblasts”\textsuperscript{83}. Despite the fact that local invasion and tumor-budding formation could partially be attributed to the paracrine involvement of EMT-promoting factors of such desmoplastic stroma\textsuperscript{110, 154}, the exact contextual milieu regulating this process remains particularly elusive and underexplored.

In an attempt to delineate the secreted profiles/signatures of the cancer invasion front, as a prerequisite to investigate for potential mediators of localized invasion and tumor budding formation, this chapter describes the generation of \textit{in-vitro} cell co-cultures consisting of colon cancer cells and cancer-associated fibroblasts, and the subsequent, comprehensive proteomic analysis of their secretomes.
2.2. EXPERIMENTAL PROCEDURES

2.2.1. Cell Culture

The human SW480, SW620, HCT116 and HT29 colon cancer, as well as 18Co normal colonic fibroblast cell lines were purchased from the American Type Culture Collection (ATCC, Rockville, MD) and were maintained in their suggested media, according to the manufacturer’s instructions. Experiments were performed within 6 passages from the initiation of cell culturing.

The following direct (cell-to-cell contact) cocultures were generated: SW480/18Co, SW620/18Co, HT29/18Co and HCT116/18Co, and were termed as: SW480Co, SW620Co, HT29Co and HCT116Co, respectively. In specific, 18Co fibroblasts were initially seeded at T-175 flasks and allowed to reach at least 50% confluence in minimum essential medium (MEM), supplemented with 10% fetal bovine serum (FBS). Then, flasks were rinsed with phosphate buffered saline (PBS) to remove dead cells. Colon cancer cells (2X10^6) were resuspended in Dulbecco’s modified Eagle’s medium (DMEM) with 10% FBS and placed on top of the pre-seeded 18Co cells. The cocultures were left to grow on DMEM, 10% FBS for at least 2-3 additional days prior to further experiments. 18Co cells showed excellent viability when cultured in either MEM or DMEM with 10% FBS, using a trypan blue assay (data not shown).

2.2.2. Secretome Analysis

Proteomic analysis of cancer cell line conditioned media (CM) (secretome analysis) was originally developed in our laboratory and was, here, performed for mono- and coculture CM (Fig 2.1.), with slight modifications, as follows: All mono- and cocultures were washed with PBS and treated with chemically-defined Chinese hamster ovary (CDCHO) medium for 2 days. Following that, all CM were collected and their volume was normalized to a total of ~1.1 μg of total protein (Coomassie colorimetric assay; Pierce biotechnology). The samples were dialyzed, using a 3.5 kDa molecular cut-off membrane (Spectrum Laboratories, Inc., CA, USA) in 4 L of 1 mM ammonium bicarbonate solution (Fisher Scientific) (with a total of 4 buffer changes). Dialyzed CM were frozen at -80 ºC and subjected to lyophilization until total dryness. The samples were then denatured by 8 M urea (Fisher), reduced with dithiothreitol (DTT) (Sigma-Aldrich) to a final concentration of 13 mM at 50 ºC for 30 min, alkylated with 500 mM iodoacetamide (Sigma-Aldrich) in the dark for 1 h at room temperature, and finally desalted using a NAP5 column (GE Healthcare), using the manufacturer’s instructions. The 1 mL eluted samples were trypsin-digested (Promega) at a molar ratio of 1:50 (trypsin:protein concentration) for 8 h. The
peptides were again lyophilized to dryness and resuspended in 120 μL of 0.26 M formic acid in 10% acetonitrile (mobile phase A).

Peptide separation was performed with two-dimensional liquid chromatography (LC), with the first dimension being strong cation exchange (SCX-LC), while the second reverse-phase (RP-LC). The samples were fractionated using an Agilent high-performance liquid chromatography (HPLC) system connected to a PolySULFOETHYL ATM column with 200-Å pore size and a diameter of 5 μm (The Nest Group Inc.). A 1-h linear gradient was used with 1 mol/L ammonium formate and 0.26 mol/L formic acid in 10% acetonitrile (mobile phase B) at a flow rate of 200 μL/ min. Fractions were collected every 5 min for the first 20 min of the run and every 2 min for the following 40 min, to a total number of 24 fractions/replicate. Of these, 16 fractions, corresponding to the highest concentration of eluted peptides, were kept for mass spectrometry. A peptide cation exchange standard (Bio-Rad), consisting of four peptides, was run at the beginning of each replicate to assess column performance. HPLC fractions were C18-extracted using a ZipTipC18 pipette tip (Millipore) and eluted in 5 μL of 90% acetonitrile, 0.1% formic acid, 10% water and 0.02% trifluoroacetic acid (Buffer B). 80 μL of 95% water, 0.1% formic acid, 5% acetonitrile and 0.02% trifluoroacetic acid (Buffer A) were added to this mixture, and 40 μL were injected via an autosampler on an Agilent HPLC. The peptides were first separated onto a 2-cm C18 guard column (inner diameter 200μm), then eluted onto a resolving 5-cm analytical C18 column (inner diameter 75 μm) with an 8-μm tip (New Objective).

The second HPLC system was coupled online to an LTQ-Orbitrap XL (Thermo Fisher Scientific) mass spectrometer, using nano-electrospray ionization (ESI) source (Proxeon Biosystems), in data-dependent mode. Each fraction was run with a 55-min gradient and eluted peptides were subjected to one full MS scan (450-1450 m/z) in the Orbitrap at a resolution of 60,000, followed by six data-dependent MS/MS scans in the linear trap quadrupole (LTQ Orbitrap). Unassigned charge states and charges 1+ and 4+ were all ignored, as depicted through the charge-state screening and preview mode. Data files were created by use of Mascot Daemon (version 2.2.0) and extract_msn. The resulting mass spectra from each fraction were analyzed using Mascot (Matrix Science; version 2.2) and X!Tandem (Global Proteome Machine Manager; version 2005.06.01) search engines, using the International Protein Index human database (IPI version 3.62, 167,894 protein sequences), which includes both forward and reverse sequences. The resulting Mascot and X!Tandem files were loaded into Scaffold (Proteome Software; version 3.6.1) to cross validate data files from both engines. Detection of a minimum of two unique peptides was required to accept positive protein identification.
FIGURE 2.1. **Proteomic pipeline for cell line secretome analysis.** Following the termination of the mono- and coculture experiments (i.e. 2-days post-seeding), supernatants are collected and centrifuged to remove dead cells. These conditioned media are subjected to three sequential steps along the pipeline: (a) sample preparation (i.e. dialysis, protein reduction alkylation and trypsin digestion), (b) peptide fractionation/separation using liquid chromatography and (c) protein identification using mass spectrometry. Proteomic datasets are then subjected to further bioinformatic analyses, depending on the question asked. CM, conditioned media; SCX, strong cation exchange; RPLC, reverse-phase liquid chromatography; MS/MS, tandem mass spectrometry.
2.2.3. Dataset Organization and Visualization

Relative quantification of proteins among the various datasets was based on spectral counting \cite{445,446}. Normalization to avoid possible operator- or machine-specific biases during the experimental part of the pipeline was performed, as previously described \cite{447}. In particular, we used the “Quantitative Value” function of Scaffold 3.6.1 software, which provides normalized spectral counts based on the total number of spectra identified in each sample. Spectrum reports were exported from Scaffold for further analysis. For dataset comparisons, VENNY (http://bioinfogp.cnb.csic.es/tools/venny/index.html) was used.

2.2.4. Secretome Dendrogram

Proteins were examined for predicted secretion with presence of signal peptide (SignalP 4.0) (http://www.cbs.dtu.dk/services/SignalP/) \cite{205,448} or without such presence (SecretomeP 2.0) (http://www.cbs.dtu.dk/services/SecretomeP) \cite{208}, and amino acids relevant to transmembrane helices (TMHMM 2.0) (http://www.cbs.dtu.dk/services/TMHMM) \cite{210}. The IPI human database v3.71 and the ENSEMBL human gene annotations (version 62) were used to map the protein identifiers to FASTA sequences, used as input to all three prediction software. The following cutoffs were considered: >0.5 (SignalP), >0.7 (SecretomeP), >0 transmembrane helices (TMHMM). An automation tool (i.e. secretome dendrogram) (Fig. 2.2.), was developed in-house in R and can be run on a Linux environment, facilitating the mapping of a list of protein identifiers (IPI number of ENSEMBL gene transcripts) to sequences, running the three prediction software mentioned above and creating a report file with consolidated predictions. The input is a list of gene names, in which case the IPI database is used to obtain the list of identifiers associated with those. For efficiency reasons, a locally installed version of SignalP and TMHMM were used to obtain predictions, whereas for SecretomeP the web-based server was queried.

2.2.5. Immunoassays

The concentration of kallikrein-6 (KLK6), -7 (KLK7) and -10 (KLK10) was measured by in-house developed sandwich-type immunoassays, as previously described \cite{449}. Concentration was determined by interpolation from a standard curve, using recombinant proteins developed in our laboratory, as standards. The concentration of MMP2 and uPA were measured with commercially-available immunoassays (Invitrogen), following the manufacturer’s instructions.
FIGURE 2.2. Secretome dendrogram. Each identified protein is sequentially checked through prediction software (i.e. TMHMM, SignalP, SecretomeP) and may move along the decision tree according to the following cutoffs for a successful pass: presence of at least one (i.e. $>0$) transmembrane helix for TMHMM, $>0.5$ score for SignalP, and $>0.7$ score for SecretomeP. The decision tree automatically assigns each protein to only one of the proposed protein groups at the end. Proteins designated as intracellular contaminants or bound to internal membranes are filtered out from the final “secretome dataset”.

2.2.6. Immunocytochemistry (ICC)

18Co cells were placed on poly-L-lysine-coated coverslips (POLY-PREP SLIDES, Sigma) at ~30,000 cells/coverslip, in MEM, 10% FBS, and were left for 24 h in Petri dishes to allow for adherence. Following this, the coverslips were washed with PBS. SW480 and SW620 cells were placed on top of the fibroblasts at approximately 20,000 cells/coverslip, in DMEM, 10% FBS and were left for 32 h to allow for adherence and proliferation. Then, the slides were washed with PBS and left for an additional 72 h, in CDCHO, to allow for paracrine signaling to occur during co-culturing. Monocultures served as controls. After 2 days, all slides were washed with PBS, fixed in 4% paraformaldehyde for 1 h and soaked in ethanol overnight. The cells were permeabilized, using 2 mL of 0.05% Triton-X-100 detergent for 5 min, then washed with PBS and covered with 0.5 mL peroxidase blocking reagent for 5 min. Cells were again washed with PBS and subsequently incubated for 1.5 h at room temperature with primary antibody, directed against alpha-SMA, diluted 1:50 in PBS. Cells were then washed with PBS and incubated with 0.5 mL goat serum blocking agent for 15 min. Cells were then washed with PBS and treated with peroxidase-conjugated secondary antibody, diluted 1:200 in PBS. Cells were then washed with PBS and incubated with 0.2 mL diaminobenzidine (DAB) chromogen buffer and counterstained with hematoxylin. Replicates of no-antibody controls were also included. Fluorescence was visualized using light microscopy (Olympus), attached to a Q-Color3 camera (Olympus) and Q-Capture Pro image analysis software.

2.2.7. Immunohistochemistry (IHC)

Paraffin-embedded tissue sections from 15 archived cases of moderately or poorly differentiated human CRC, showing an intense desmoplastic reaction in invasive areas were kindly provided by Dr. Constantina Petraki. The immunohistochemical staining was performed using the Bond automated immunohistochemistry system (Bondmax, Leica Microsystems, UK)-pretreatment with epitope retrieval (pH 8). Mouse monoclonal antibodies against the following proteins were used: Cytokeratin cocktail (Cell Marque, clone A1E&AE3, 1:70), alpha-SMA (DAKO, clone 1A4, 1:100), Collagen type III (Biogenex, clone HWD1, 1:100) and Collagen type XII (Sigma-Aldrich, clone 1303, 1:30).

2.2.8. Statistical analysis

All data are presented as bars of means with their standard error.
2.3. RESULTS

2.3.1. Development of an *in-vitro* Coculture System, Mimicking the Paracrine Interactions in the Colorectal Cancer Desmoplastic Microenvironment

Among the plethora of colon cancer cell lines, we selected the SW480/SW620 as a reasonable *in vitro* system for our initial experiments. This system allows us to capture the heterogeneity of the disease to some extent, since the SW480/SW620 cell lines were obtained from the same patient, but at a different tumor stage. We developed cell contact, two-dimensional cocultures of SW480/SW620 and 18Co cells (SW480Co and SW620Co, respectively) and used all the relevant monocultures as controls (Fig. 3.3A.). Similar viable cocultures were generated with two additional colon cancer cell lines, the HT29 and HCT116 (Fig. 3.3A.), suggesting that the ability of 18Co cells to form healthy cocultures with colon cancer cell lines is not cell line-specific.

Given that paracrine signaling between normal colonic fibroblasts and cancer cells is present in coculture conditions, we hypothesized that 18Co should be transdifferentiated into CAFs *in-vitro*, an observation that would be in agreement with the literature, and would render our coculture system as an accurate representation of the desmoplastic invasion front. When we used specific immunoassays to measure MMP2 and uPA, two prominent markers of CAFs, in mono- and coculture CM, we found minimal secretion in 18Co monocultures and no secretion in SW480/SW620 cell lines (Fig. 2.3B.). However, their secretion was aberrantly increased in both SW480Co and SW620Co cocultures (Fig. 2.3B.). Similar results were reproduced in the two additional cocultures, HT29Co and HCT116Co (Fig. 2.3B.).

Using ICC, we also tested the expression of alpha-SMA, an intermediate filament protein, which serves as marker of CAFs and the myofibroblastic phenotype in general. Although normal fibroblasts might occasionally express low levels of alpha-SMA, we found that alpha-SMA expression was higher in the SW480Co/SW620Co fibroblasts, rather than in monocultures of 18Co cells (Fig. 2.3C.).

Together, these phenotypic observations demonstrate that *in vitro* interactions of cancer cells with fibroblasts, may lead to the differential expression of prominent markers of CAFs. Thus, our proposed SW480Co/SW620Co coculture model could provide an *in vitro* representation of the tumor invasion front and the desmoplastic microenvironment.
FIGURE 2.3. Development of in-vitro cocultures, mimicking the CRC desmoplastic microenvironment. (A) Morphological assessment of in-vitro mono- and cocultures. The phenotype of normal colonic fibroblasts (18Co) and colon cancer cell lines (SW480, SW620, HT29, HCT116) is shown in the first row, while that of the respective cocultures is shown in the second row. The cocultures are named as SW480Co, SW620Co, HT29Co and HCT116Co, respectively. (B) MMP2 and uPA secretion in CM from all mono- and cocultures (ELISA). Note that the increased secretion of both proteins in the coculture conditions. Bars demonstrate means with standard error. (C) alpha-SMA immunocytochemistry in mono- and cocultures. Note that in contrast to monocytes, the 18Co cells in coculture conditions express alpha-SMA. Cancer cells are negative for alpha-SMA. All magnifications x40.
2.3.2. Comprehensive Proteomic Analysis of Mono- and Coculture Conditioned Media

We collected CM from all mono- and cocultures after a 2-day incubation period and subjected them to comprehensive secretome analysis, following an *in-house* protocol with slight modifications, as explained in Experimental Procedures and Fig. 2.1. We used the stringent criterion of a minimum of two peptide hits for protein identification and all experiments were performed in triplicate to increase the accuracy of identification. Therefore, a total of 15 CM samples, corresponding to five experimental conditions (SW480, SW620, 18Co, SW480Co and SW620Co) were collected, trypsin-digested and subjected to SCX-RPLC-MS/MS. We totally identified 1393, 1213, 423, 1283 and 1033 proteins for SW480, SW620, 18Co, SW480Co and SW620Co, respectively, with high reproducibility among the three replicates (ranging from 63% to 93%) (Fig. 2.4.) A total of 2142 non-redundant proteins were identified with the combination of all protein datasets. To our knowledge, there is no other proteomic study to date that has delineated the secreted proteome of the 18Co normal colonic fibroblast cell line and consequently of its respective cocultures.
Figure 2.4, Secretome analysis. Proteomic analysis of mono-culture and coculture CM. Venn diagrams show reproducibility among triplicates. In squares are the numbers of identified proteins with a minimum of two peptide hits. Monocultures are depicted with red and cocultures with blue color. The monoculture datasets have been used as exclusion lists to the coculture ones, identifying a dataset of 286 coculture-specific proteins (depicted with an asterisk).
2.3.3. Robustness and Quality of Mass Spectrometry-based Protein Identification using a Panel of Internal Control Proteins

To test the accuracy of the MS data, we considered a small panel of four secreted internal control proteins, including KLK6, KLK7, KLK10 and MMP2. Using specific immunoassays previously developed in our laboratory\(^{449}\) for the former three proteins, and a commercially-available immunoassay for the latter, we determined their expression levels in all CM subjected to secretome analysis, and subsequently associated them with spectral counting (i.e. independent MS/MS events). In general, we observed that MS-based identification and label-free quantification were robust enough and showed similar trends of expression with all four specific ELISAs (Fig. 2.5.). In brief, KLK6 was found in all but the 18Co secretome and was identified with >10 spectral counts in SW480, SW620 and their cocultures. KLK7 was identified with 3 spectral counts in the SW480 secretome; no positive identification was made in SW480Co secretome. Accordingly, KLK7 immunoassay verified this result, as KLK7 was detectable in low amounts (<2.5 ug/L), only in the SW480 CM. KLK10 was found in SW480 and SW480Co secretomes using the specific immunoassay and was identified with >7 spectral counts in the LC-MS/MS analysis. Finally, MMP2 was secreted in large amounts by 18Co normal fibroblasts (~40 ug/L at day 4), and the relevant cocultures, but not in monocultures of cancer cells. MMP2 was identified with a large number of spectral counts (~25- 50) in both 18Co cells and the cocultures. Collectively, the expression trends of the internal control proteins do not contradict between label-free quantification and immunoassays, and further suggest that KLKs may be secreted from the epithelial compartment of tumors, while MMP2 is mainly stromatogenic.
FIGURE 2.5. Robustness and quality control of mass spectrometry-based protein identification. A panel of four internal control proteins (KLK6, KLK7, KLK10 and MMP2) was considered and tested for correlation of MS/MS identification (upper panel) with respective immunoassays (lower graphs). Regarding spectral counting, numbers in the upper brackets demonstrate unique peptides, while numbers in lower brackets demonstrate unique spectra for each replicate independently. Regarding the immunoassays, all bars in the graphs represent mean values (ug/L) with standard error at either 2 (black) or 4 (grey) day supernatant for the respective culture.
2.3.4. Delineating the Desmoplastic Invasion Front Exoproteome

Combination of the two coculture datasets (SW480Co and SW620Co) gave rise to 1551 non-redundant proteins (Fig. 2.4.). In an attempt to generate a comprehensive, heterotypic dataset of secreted candidate proteins (i.e. exoproteome), exclusively present in the desmoplastic invasion front, we first used all monoculture datasets (SW480, SW620, 18Co) as exclusion datasets, to retain the “coculture-specific” proteins. This filter resulted in 286 proteins (Fig. 2.4.), which were either *de novo* expressed in the cocultures, or could be expressed in monocultures in small amounts, not detectable via LC-MS/MS. This classification also gave rise to 591 monoculture-specific proteins and 1265 common proteins (i.e. present in at least one mono- and one coculture dataset) as shown in the Venn diagrams (Fig. 2.4.). Although these 591 monoculture-specific proteins could also be of functional importance, we decided to focus on the 286-protein dataset, since they were the most promising candidates. We cannot exclude the possibility that some of these proteins (probably of low-abundance) were not detected in some monoculture CM, due to methodological and instrumental limitations (relatively low sensitivity) and thus are “false-positives”. However, the three biological replicates and the 2-peptide hit criterion should minimize such biases.

Secondly, we filtered the 286 protein dataset through our proposed “secretome dendrogram” (Fig. 2.2.; for details, refer to Experimental Procedures) to remove all intracellular contaminants and internal membrane-bound proteins and focus on the exoproteome of the desmoplastic invasion front. With this filter we identified 152 proteins. Of these, 98 were predicted as non-membranous and 54 as membranous, spanning a number of transmembrane helices from 1 to 10. Of the non-membranous proteins, 75 were predicted as secreted with the classical secretory pathway and 23 with the non-classical one. Since this 152-protein dataset represents a dataset of secreted proteins originating from tumor host cell interactions, we assigned it with the term “Desmoplastic Protein Dataset” (DPD) (Appendix; Table A).

2.3.5. Collagen Type XII, a Novel Marker of Myofibroblastic Differentiation in CRC

Having completed verification experiments (chapter 2.3.3.), which demonstrated the integrity of the MS-based protein identification, as well as prioritization criteria, which defined DPD as a novel molecular signature of the cancer invasion front (chapter 2.3.4.), we sought to investigate whether the constituents of such signature hold any translational importance. Given that our approach for determining the signature was based on an *in-vitro* setting, we proceeded to establish proof-of-concept evidence before subjecting the DPD signature to sophisticated and complicated pathway analysis.
Of all proteins, we decided to focus on collagen type XII (COL12A1) with the following rationale: First, data obtained from our proteomic analysis suggests that it is extremely unlikely that COL12A1 could be a false positive protein (i.e. it was identified with 68 unique peptides in the SW480Co and with 92 unique peptides in the SW620Co CM. Second, various ECM components deposited in desmoplastic lesions, especially collagens (i.e. collagen types I and III), have been widely demonstrated to be CAF markers. Third, COL12A1 has been previously linked to fibrotic diseases, but with non-cancerous pathology. Fourth, COL12A1 has been previously demonstrated (by proteomics) to be expressed by smooth muscle cells perivascular areas. In this global proteomic approach, a mass spectrometry-based platform for identification of vascular extracellular proteins in human aorta samples, revealed a novel ECM component signature, including various proteoglycans and collagens (such as collagen type XII). Therefore, there is intriguing evidence that COL12A1 is being expressed in certain diseases with similar pathology to desmoplasia (i.e. fibrotic diseases) by certain cell types with similar phenotype to CAFs (i.e. smooth muscle cells, myofibroblasts).

To address our original question, we first defined the cancerous and stromal subpopulations at invasive fronts of desmoplastic lesions in a small cohort (N=15) of CRC patients (Fig. 2.6A.). In such areas, the stromal reaction mostly consisted of collagen, CAFs and inflammatory cells. Cancer cells invading the stroma were characterized by strong keratin expression. A strong cytoplasmic alpha-SMA IHC expression was observed in CAFs around CRC invasive areas. These are the areas of paracrine interaction that were, in fact, recapitulated through our *in-vitro* coculture system, and which were found eligible for evaluating COL12A1 expression. Collagen type III expression was used as quality control to verify the presence of desmoplastic microenvironment throughout. Both collagen type III and XII were strongly co-expressed in the collagenous matrix around tumor cell cohorts, as well as in CAFs (black arrows; Fig.2.6B). Thus, our data strongly support the notion that collagen type XII could be an indicator of myofibroblastic induction in CRC desmoplastic lesions.

Interestingly, small clusters of cancer cells positioned at the interaction line demonstrated strong immunoeexpression for COL12A1 (arrowheads; Fig. 2.6B.), while proximal tumor cells (positioned in the core of the cohort) were negative. This was not the case for collagen type III, since this type of collagen was characteristic of the stromal compartment, exclusively (Fig. 2.6B.). This might suggest that paracrine interactions between cancer cells and CAFs are sufficient for COL12A1 expression in both cell types, an observation that renders putative translational importance and proof-of-concept that DPD could be enriched in proteins that expressed in the CRC desmoplastic invasion front.
FIGURE 2.6. Immunohistochemical determination of invasion front and association with collagen type XII. (A) Immunohistochemical markers defining the cell populations of the CRC desmoplastic invasion front. Tumor invasion front area (red arrows) of colorectal carcinoma (black arrows) with myofibroblastic stromal reaction (green arrows) containing tumor budding cells (blue arrows). (i and iv) H-E stain. (ii and v) Pan-keratin immunoexpressed strongly in CRC cells and slightly in myofibroblasts. (iii and vi) alpha-SMA immunoexpressed in myofibroblastic stromal reaction. Magnifications i, ii and iii x 200, and iv, v and vi x400. (B) Immunohistochemical profile of collagens type III and XII in CRC desmoplastic invasion fronts. (i-iii) Collagen type III immunohistochemical expression in the stromal collagen and myofibroblasts (arrows) around the tumor invasion front in cases of CRC. (iv-vi) Collagen type XII immunohistochemical expression in the stromal collagen and myofibroblasts (arrows) around the tumor invasion front in the same cases of CRC, respectively, and in tumor cells of the tumor invasion front (arrowheads). Magnifications i, iv and vi x200, and ii, iii and v x400.
2.4. DISCUSSION

2.4.1. Innovation and Critical Evaluation of the Proteomic Approach

A significant innovation in the experiments presented here is that they together comprise one of the very few attempts to characterize secreted proteins of the “cancer invasion front” in a high-throughput fashion, and the very first one related to CRC, in particular. Previous high-throughput attempts have mainly focused on delineating transcriptomic profiles (i.e. microarrays) of the desmoplastic microenvironment, probably because proteomic technologies and mass spectrometers were not very sensitive and the experimental pipelines were not particularly concrete. In the modern era of personalized medicine though, this landscape has significantly changed, and the field of cancer secretomics has become more robust and offers reproducible results. This could prove to be a very important contribution in cancer research, because the co-assessment of transcriptome and proteome profiling is more powerful in elucidating disease mechanism compared to gaining information from either of the two alone.

So far, there is a considerable bulk of publicly-available databases, which serve as repositories for thousands of transcriptome experiments. Interestingly, databases focusing on the proteome of the cells had been lacking for a long time, and only during the last decade, there has been a massive outrush of proteomic information that has been organized into similar repositories. Our vision for the future is the integration of highly-sophisticated statistical methodology for further mining and meta-analysis of this vast information, to ultimately come to a clear understanding of their pathobiological importance and, of course, to suggest appropriate clinical manifestations.

We allege that this particular study might have been prone to a certain number of technical biases and mistakes, the most prominent of which are listed below.

1. Our inability to distinguish whether a protein of the DPD originates from the cancer cells themselves or the CAFs in the coculture conditions. If we wish to gain such information subsequent experiments need to be performed, as the coculture secretomes are apparently far more complicated in content.

2. Since proteins of the DPD were not identified via MS analysis in any of the control monocultures, their presence in the coculture CM could be considered as de novo, attributed to the paracrine tumor host-cell interactions. However this de novo speculation should be vigilantly interpreted, since it is possible that low-abundance proteins (e.g. cytokines) might not have been identified in monocultures, due to relatively low method sensitivity.
Our in vitro model system involved only two colon cancer cell lines (SW480 and SW620), which were derived from a different stage of the same patient. Thus, the proteins of the DPD may have been influenced by the genetic background of these particular cell lines. For instance, it is known that SW480 is deficient in SMAD4, although it is responsive to non-canonical TGF-beta signaling. This SMAD4-deficient molecular portrait, and the exoproteome, for which it could be responsible for, could theoretically only reflect to certain subgroups of CRC patients. This problem could be resolved with the inclusion of many different cell lines to recapitulate the entire heterogeneity of the disease, although this also entails additional cost and time.

For the above mentioned reasons, we foresee that this study is mostly valuable for proof-of-concept and methodological aspects. For the latter, an integration of multiple sources should always be performed to maximize chances for success. Overall, the DPD should be evaluated with great caution, as it represents only a context-dependent, proteomic snapshot of the dynamic interactions of the desmoplastic invasion front.

2.4.2. A Deeper Look into the Desmoplastic Protein Dataset

Notably, this proteomic strategy allowed us to re-identify previously-reported markers of myofibroblastic differentiation, such as the secreted proteoglycan versican (VCAN) isoforms V0 and V1, urokinase-type plasminogen activator (uPA), as well as the motogenic cytokine, hepatocyte growth factor (HGF), among others (Appendix; Table A). These observations are highly consistent and reminiscent of CAF activity and form a paradigm that our proteomic strategy and model system may be particularly efficient in recapitulating the molecular interactions of the CRC desmoplastic interface.

However, the absence of yet other reported markers of myofibroblastic induction, such as MMP2 and MMP9, could be easily observed in DPD. This could draw certain concerns since these markers are very prominent during the transition of NFs into CAFs, and, as such, it seems that DPD fails to efficiently recapitulate the desmoplastic interface in this aspect. Nevertheless, it should be mentioned that the DPD only includes those proteins that were de novo secreted in the coculture conditions. It is very possible that many other proteins that are upregulated in CAFs, but are also secreted by NFs, and thus present in the monoculture secretomes, might have skipped our attention during the initial filtering (Fig. 2.3.). To overcome this issue, we should develop a semi-quantitative approach, by also including proteins identified in all, but differentially expressed between mono- and co-culture conditions. However, this is not possible in the current setting, as proteins cannot be easily normalized for quantification in CM. Of note, both MMP2 and MMP9 were identified with a prominent number of spectral counts in the coculture...
conditions, but were also identified in 18Co monocultures (data not shown; refer to our published paper\textsuperscript{453} for raw data), and this is the reason why they were not further prioritized and included in the DPD.

Remarkably, a thorough examination of the DPD reveals the co-existence of positive and negative regulators of carcinogenesis. Despite paradoxical, it is indeed quite obscure whether the desmoplastic microenvironment plays a tumor-promoting or tumor-suppressive role in colorectal and other cancers, as both sides have been supported with robust scientific evidence\textsuperscript{83,111}. The current paradigm states that the desmoplastic stroma is a source of various factors that may either promote or suppress cancer progression, and it is actually the thin balance between them -along with contextual information presented from other stromal interactions- the principal determinant that will dictate the fate of the proximal cancer cells\textsuperscript{57,83}. On one hand, the desmoplastic reaction represents an effort of the healthy tissue to entrap and encircle the tumor, by providing a biological limitation to the expansion of cancer collectives. On the other side, cancer cells seem to adapt with phenomenal plasticity in this continuously modified landscape and to develop alternative molecular pathways to shift desmoplastic proteins towards their own favor and acquire a series of essential hallmarks for achieving metastasis\textsuperscript{462-465}.

2.4.3. The Expression of Collagen type XII in the CRC Desmoplastic Invasion Front

Collagen type XII was the most abundantly-expressed protein among the 152 candidates of the DPD signature, and quite interestingly, no evidence linking it to CAF expression currently exists. Collagen type XII belongs to the fibril-associated collagens with interrupted triple helices (FACITs), which are, in fact, fibrillar collagen organizers. Collagen types XII and XIV both have one domain that anchors the molecule to the surface of another fibril, such for instance to collagens type I or III, and three “finger-like” domains\textsuperscript{466,467}. By residing in collagen I fibrils, collagen type XII participates in fibril structure interaction and organization, by further stabilizing them\textsuperscript{468}. Thus, our proteomic investigations showed that collagen type XII could be a product of paracrine interaction in a heterotypic tumor microenvironment, an observation which suggests that CAFs may possibly attempt to reorganize the collagen around the cancer invasion front, through FACIT secretion. Additionally our \textit{in vitro} experiments revealed that collagen type XII was abundantly secreted in both SW480Co and SW620Co cocultures, but we could not discern which cell type was responsible for this production. Interestingly, the immunohistochemical analysis revealed its expression by cancer cells lining the desmoplastic front and by cancer cells of the tumor budding subpopulation, in addition to CAFs. It is quite interesting that Collagen type XII, in contrast to Collagen type III, was also expressed by the epithelial compartment of the tumor and not solely by the desmoplastic stroma. We, therefore, speculate that Collagen type XII could also serve as a marker of dedifferentiation.
(i.e. EMT), since we have extensively explained (chapter 1.1.4) that cancer cells are progressively losing their epithelial phenotype towards the desmoplastic interface 154, 469, 470.

To better understand the reasons behind such aberrant FACIT secretion in the desmoplastic interface, it is necessary to first introduce the concept of ‘matrix stiffness’ and demonstrate its causative implication in cancer development and progression 64. Indeed, the development of many cancers is accompanied by progressive sclerosis (or stiffening) of peritumoral ECM. For instance, mammographic density measurements in human patients shows that mammary tumor tissue and tumor-adjacent stroma are 5 to 20 times harder than the normal mammary gland 471. The enhancement of tumor stroma durability is regulated by CAFs and allows tumors to overcome the increased mechanic pressure 472. Accumulating evidence supports a generic hypothesis that the CAF population enforces the persistent growth and metastatic potential of the cancer cells, allowing them to overcome the topographic restrictions due to the desmoplastic microenvironment 64. Collagen is the most abundant ECM scaffolding protein in the stroma and the main contributor of tissue tensile strength. Its metabolism is aberrantly deregulated during development of cancer, where increased collagen expression, deposition, remodeling, and organization are deployed to support the neoplastic tissue 473. Fibroblasts, along with other recruited stromal cells such as immune and endothelial cells, elaborate a plethora of growth factors, cytokines, and chemokines to support this 472. It has been suggested that collagen cross-linking in tumor boundaries, could increase ECM density and interstitial pressure for the confinement of the tumor and the preservation of tissue homeostasis. CAFs seem to be the key elements participating in collagen cross-linking, resulting in remodeled and reoriented collagen in fibrotic lesions, collectively known by the term ‘cancer-associated collagen’ 472-476.

A widely explored mechanism by which CAFs achieve collagen cross-linking is via production of lysyl-6-oxidase (LOX), a copper-dependent amine oxidase, which initiates the process of covalent intra- and intermolecular cross-linking by oxidatively deaminating specific lysine and hydroxylysine residues located in telopeptide domains 477, 478. Using rheological measurements and second harmonics generation imaging, it was shown that mammary glands conditioned with LOX-expressing fibroblasts were harder in consistency than those containing LOX-negative fibroblasts. LOX-expressing fibroblasts enhanced the deposition of fibrillar and linearized (cross-linked) collagen 477. Presumably, LOX is a direct product of the desmoplastic reaction and its expression occurs at the tumor-host cell interface. LOX can be induced by both fibrogenic pathways such as TGF-beta, as well as by hypoxia, which often occurs during cancer progression 479, 480. Whether collagen type XII holds an important and relevant role to that of the well-studied LOX, and whether these two molecules co-operate or act independently to mediate collagen cross-linking are subjects that remain to be elucidated in the future 64.
We believe that collagen cross-linking elicited by CAFs may be hiding a very important evolutionary defensive mechanism, which is hardwired into the fibroblast progenitor genome and “awakens” as a memory in the CAF population. This working hypothesis supports the following correlation. In healthy non-neoplastic tissues, this mechanism is exhibited by activated fibroblasts/myofibroblasts, in order to protect healthy tissues from potential mechanical damage, by linearizing and stabilizing the collagen framework. In the tumor context, the CAFs retain parts of this memory to restrict the expansion of the cancer compartment. However, the tumor is using this defensive mechanism for its own advantage; by linearizing the disoriented collagen that is abundantly produced in desmoplastic interfaces, it creates a more stable microenvironment, which can now be used for further growth and expansion while at the same time it might protect from entrance of toxic substances, including chemotherapeutic drugs.

2.4.4. Conclusive Remarks

In the first chapter of this thesis, we have applied a mass spectrometry-based proteomic strategy to investigate tumor-host cell interactions, based on proteomic analysis of CM (secretome analysis), which were derived from CRC cell line monocultures or their co-cultures with NFs. After verifying that this model system may, in part, mimic molecular features of the colorectal cancer invasion front, we subsequently used bioinformatics tools to propose the secreted molecular signature of this interface. We anticipate that this 152-candidate signature, which we termed “desmoplastic protein dataset” (DPD) for purposes of designating biological relevance to it, may enclose regulatory factors/mediators of local invasion and tumor-budding formation.
CHAPTER 3
Pathway Profiling of the Colorectal Cancer Desmoplastic Invasion Front

Sections of this chapter have been published in Molecular Oncology:

Karagiannis GS, Berk A, Dimitromanolakis A, Diamandis EP. Enrichment Map Profiling of the Cancer Invasion Front Suggests Regulation of Colorectal Cancer Progression by Bone Morphogenetic Protein Antagonist, Gremlin-1. Molecular Oncology, 2013; 7: 826-839.

Copyright permission has been granted.
3.1. INTRODUCTION

In an attempt to decipher potential factors that could serve as an appropriate context for local invasion and tumor budding formation in the desmoplastic invasion front, we have previously developed a comprehensive proteomic signature, consisting of 152-secreted candidates, identified through proteomic delineation of colon cancer/cancer associated fibroblast co-cultures. Given that our proteomic signature might reveal potential mediators of tumor progression at the level of the cancer invasion front, it is conceivable that a sophisticated method of unraveling overrepresented pathways would be essential to prioritize the most promising candidates. Currently, modern bioinformatics tools and software for molecular pathway analysis have been of great assistance to organize such datasets, generated from high-throughput genomic/proteomic pipelines, as in our case.

In an attempt to mine rationally into the desmoplastic protein dataset and elucidate mediators of local invasion and tumor budding formation, this chapter describes the deployment of a two-layered, integrative bioinformatics strategy that uses independent molecular algorithms, aiming at the organization of the proteomic dataset into useful pathways as predominant principle.
3.2. EXPERIMENTAL PROCEDURES

3.2.1. Experimental Design

Our integrative approach involves two subsequent steps, using a variety of molecular algorithms to achieve pathway analysis (Fig. 3.1.). The first level involves investigation for potential enrichment of the desmoplastic protein dataset in molecular concepts of interest\(^{482}\), by performing Gene Set Enrichment Map analysis with gene ontology annotations for biological process\(^ {483}\). Overrepresentation of specific themes will guide us to select those that are of interest for our hypothesis, and these will be subsequently brought to the second level of analysis. This involves the identification of molecular hubs that participate in the organization of these themes\(^ {484}\), by using protein-protein interaction network analysis, like, for example, STRING analysis\(^ {485}\).

3.2.2. Retrieval of Proteomic Dataset

Bioinformatic analyses were performed on the desmoplastic protein dataset (DPD), a proteomic dataset described on Chapter 2, and listed on Table A (Appendix).

3.2.3. Performance and Visualization of Enrichment Analysis

Enrichment map profiling was performed as previously described\(^ {486}\). In brief, BINGO (v.2.44) was utilized to calculate over-representation of Gene Ontology (GO) “biological process” terms among the input gene list (DPD; Appendix; Table A). The hypergeometric test was performed to assess the significance of the enrichment and resulting p-values were FDR-corrected using the Benjamini & Hochberg method (p < 0.05). Data from the enriched GO annotations was exported. Functional enrichment visualization was constructed by the enrichment map plug-in in Cytoscape\(^ {483, 487}\), using the exported BINGO results as the input. The Jaccard coefficient was used at a cutoff of 0.5 to connect related GO biological process terms and create an enrichment network. Parameters for the selection of the enrichment results, which appear on the network, were set to a p-value cutoff of 0.005 and FDR < 10%. The generated clusters within the map were named after a commercially-available word cloud algorithm for word frequencies (Wordle) (http://www.wordle.net/).
3.2.4. Pathway Analysis using Ingenuity Pathway Analysis (IPA)

Pathway analysis was performed with ingenuity pathway analysis (Ingenuity Systems; IPA) software (http://www.ingenuity.com/), as previously described \(^{488,489}\). In brief, the protein dataset of interest (i.e. DPD) was uploaded into the application as standard human gene symbols. Each gene identifier was mapped to its corresponding gene object in the ingenuity pathways knowledge base (IPKB). The IPKB, containing a large network of curated molecular interactions and pathways, was searched to find sub-networks enriched in genes of interest. Graphical representations of these sub-networks, containing direct and indirect molecular relationships, were generated. Genes or proteins are illustrated as nodes and molecular relationships as connecting lines between two nodes (direct relationships as normal lines; indirect relationships as dashed lines). Molecular relationships are supported by at least one literature reference, or by canonical information stored in the IPKB. Grey nodes represent genes of interest, while white nodes represent hubs that were added by the IPA algorithm to connect a set of genes of interest.

3.2.5. Generation of Protein-Protein Interaction Networks using STRING

Protein-protein interaction analysis was performed using ‘search tool for retrieval of interacting genes/proteins’ (STRING) (http://string-db.org/) \(^{485,490}\). The protein dataset of interest (i.e. DPD) was uploaded into the application using gene identifiers and complete lists of human orthologs were included in the network visualization. Protein-protein interaction networks were algorithmically created requesting high confidence scores (>0.7). Proteins are indicated with nodes and protein interactions with connecting lines. Confidence views were requested for the visualization.
FIGURE 3.1. Two-layered bioinformatics strategy for the organization of the desmoplastic protein dataset. Briefly, the first layer involves the identification of enriched molecular concepts of interest in our dataset, while the second layer involves the identification of certain molecular hubs/drivers within selected molecular concepts from the first layer.
3.3. RESULTS

3.3.1. Overview of Enrichment Map Profiling of the DPD

We have previously generated in-vitro cocultures of colon cancer cells and normal colonic fibroblasts (SW480Co/SW620Co), to mimic the desmoplastic tumor-host cell interface in CRC. By performing proteomic analysis of coculture CM, we proposed the secreted 152-protein DPD signature. Since desmoplasia affects cancer development and progression, we reasoned that DPD might hold key factors, regulating malignancy at CRC cancer invasion fronts. To test this, we unraveled overrepresented themes in DPD through enrichment analysis in Gene Ontology (GO) annotations for biological process. This analysis resulted in the significant (p<0.05) overrepresentation of 155 GO terms. Since several annotations are branched together, we visualized the analysis as an enrichment network, which algorithmically clustered GO terms with highly similar content, using the enrichment map plug-in in cytoscape environment. To better define the nature of these clusters, we used a word-frequency detection algorithm to automatically generate prevailing keywords in these clusters. The resulting keywords corresponded to well-established hallmarks of cancer, namely “cell proliferation”, “cell migration”, “cell motility”, “cell adhesion”, “neoangiogenesis”, “neovascularization”, “development and morphogenesis”, “inflammation” among others (Fig. 3.2.). Further observations on the nature of these clusters demonstrated the presence of distinct molecular signatures in the enrichment map, whose relevance with the desmoplastic microenvironment and myofibroblastic differentiation/function alleged a positive proof-of-concept for our original proteomic experiment. A few such instances are the identification of the “collagen organization and assembly”, as well as “wound-healing” clusters. Indeed, the postulation of cancers as wounds than never heal has been done long time ago, and it is now well-known that especially the desmoplastic microenvironment hosts molecular phenomena with profound similarities with those observed during wound healing.
FIGURE 3.2. Enrichment map profiling of the desmoplastic protein dataset. The network reveals overrepresentation of functional clusters associated with cancer development and progression. Nodes represent GO terms and lines demonstrate their connectivity. Dashed lines encircle groups of relevant GO terms into functional clusters, named from a commercially-available word cloud algorithm.
3.3.2. Bone Morphogenetic Protein Antagonist/Inhibitor Signature is Overrepresented in DPD

In an effort to characterize specific signaling/metabolic pathways with causative regulation of neoplastic progression at the level of the cancer invasion front, we observed the TGF-beta superfamily of proteins harbored one major regulatory network in the enrichment map (Fig. 3.2.). Three clusters each one corresponding to one of the three documented families of the TGF-b-superfamily of proteins (i.e. the TGF-beta pathway, the activin/inhibin pathway and the BMP pathway) were all overrepresented in the enrichment map profiling of DPD (Fig. 3.2.).

The corresponding GO terms characterizing the three family clusters were associated with the regulation of signal transduction, as well as with various developmental processes, such as EMT and cell differentiation (Appendix; Fig. A). This observation was very promising to fulfill aspects of our hypothesis, especially since our rationale proposes that various regulators of cancer cell adhesion, motility and (de)differentiation (which may be present in the desmoplastic invasion front) could be rendered responsible for mediating localized invasion, tumor-budding formation and further progression of the metastatic cascade.

On this basis, we further looked into the constituents of these clusters, and found that they comprised of either ligands/receptors (i.e. TGFB2, TGFBR3, GDF6, INHBA) that participate in signal propagation (group I), or regulatory proteins (i.e. FST, FSTL3, HTRA3, GREM1, LTBP1) that antagonize signal transduction (group II) (Appendix; Fig. B). As suggested by two independent pathway analyses (IPA and STRING), which use diverse algorithms for generating protein-protein interaction networks, both group I and group II members of the TGF-beta-superfamily cluster were found under a common network, nucleated around representative components/hubs of the TGF-beta-pathway (Figs. 3.3A. and 3.3B.). Of note though, group I proteins exclusively belonged to the TGF-beta and activin/inhibin pathways and none belonged to the BMP pathway. On the contrary, group II proteins almost exclusively belonged to the BMP pathway, and comprised of the BMP antagonists gremlin-1 (GREM1), follistatin (FST), follistatin-like 3 (FSTL3) and high-temperature requirement A3 (HTRA3). Consequently, our data suggest that cancer cell-CAF interactions preferentially disrupt the BMP pathway propagation at the desmoplastic invasion front, through secretion of multiple bone morphogenetic protein inhibitors (BMPIs).
FIGURE 3.3. Pathway analysis for identification of molecular hubs. The genes corresponding to the TGF-beta-superfamily cluster as the input list. Both networks demonstrate that all these genes are grouped together and some share the same downstream effectors. (A) IPA network. (B) STRING network. Grey (IPA) and colored (STRING) nodes correspond to members of the TGF-beta-superfamily of proteins found in DPD, while white nodes represent other members of the TGF-b pathway that interconnect to the network.
3.4. DISCUSSION

3.4.1. An In-Depth Look into GSEA Analysis Indicates ‘Myofibroblastic’ Signatures in DPD

As explained above, GSEA and enrichment map analysis have indicated the presence of clusters comprising of molecular concepts that demonstrate the myofibroblastic/desmoplastic nature of the tested dataset. These clusters can be categorized as follows: (I) “wound healing” cluster, (II) clusters associated with fibroblast-to-myofibroblast phenotype switching, (III) clusters associated with regulation of stromal responses, such as angiogenesis and inflammation and, (IV) clusters regulating cell-to-matrix adhesion (i.e. integrin signaling), reminiscent of those in desmoplastic CIFs. It should be made quite clear that, in this thesis, we have focused and examined the tumorigenic implications of the bone morphogenetic protein inhibitor group of proteins, which comprise only a small part of the entire network of deregulated pathways, and that future efforts should encompass in hypothesis generation and future investigations.

The recruitment of CAFs by cancer cells has been paralleled with the transdifferentiation of resident, quiescent fibroblasts into myofibroblasts in tissues undergoing wound healing. Consistent with this, a “wound healing” cluster was determined in the enrichment map profiling of DPD (Fig. 3.2.). This cluster was characterized by two corresponding GO annotations, “response to wounding” and “wound healing”, in close proximity to GO annotations, associated with immune response and inflammation (Fig. 3.2.). It was also linked to the function of proteins, such as members of the complement system (e.g. C4A, C4B, CFD and CFH), which have been previously linked to both wound healing and cancer progression.

The myofibroblastic phenotype of CAFs is characterized by increased proliferation, migration, contractility, ECM remodeling and collagen organization/assembly, all of which are regulated from a de novo deployed smooth muscle-like gene and protein expression machinery. In a supportive fashion, associated clusters, such as “smooth muscle migration”, “collagen organization and assembly” and “ECM-glycosaminoglycan organization” were found in the enrichment map (Fig 3.2.). Of note, the corresponding GO annotations demarcated various ECM-remodeling proteins, including lysyl-6-oxidase (LOX) and COL12A1, two typical collagen-cross linkers, which are associated with regulation of matrix stiffness and mechanotransduction of the peritumoral desmoplastic stroma, and interestingly, we have previously demonstrated COL12A1 expression in alpha-SMA positive CAFs, as well as in cancer cells lining the desmoplastic CIFs and tumor budding cells in CRC patients.

Desmoplasia has been generally considered as a positive regulator of the angiogenic switch, with the exception of a few cancer types, such as pancreatic ductal adenocarcinoma. In CRC, colonic CAFs may locally promote angiogenesis, an observation supported by the enrichment map profiling of DPD.
whereby two relevant clusters, “neoangiogenesis” and “neovascularization”, were identified (Fig. 3.2.). Moreover, an “inflammation” cluster was identified, (Fig. 3.2.) further verifying accumulative literature that desmoplasia participates in reciprocal signaling with the inflammatory microenvironment. Members of the integrin family and focal adhesion machinery, such as ITGA2 and ITGA5, were found enriched in DPD. This was further shown through the identification of an “integrin pathway” cluster with respective GO terms related to the integrin signaling pathway in the enrichment map analysis (Fig. 3.2.). These observations verify collective literature, which supports the de novo deployment of an altered cell-to-matrix biological program and focal adhesion machinery in desmoplastic CIFs, which further supports and drives tumorigenesis.

3.4.2. The Importance behind Integrating Multiple Bioinformatics Platforms and Tools

Currently, pathway analysis software seems to be a major strategy for conceptualizing molecular pathways of interest in large proteomic datasets, especially those derived from mass spectrometry-based experiments. Due to the high load of information in such datasets, here, we performed an unbiased integrative pathway dissection of our previously described DPD, using three independent pathway analysis tools: enrichment map profiling, IPA and STRING. The enrichment map profiling pointed to the TGF-beta superfamily pathway, as an enriched canonical pathway in our dataset (1st layer of analysis; Fig. 3.2.), whereas the remaining two pathway analyses, Ingenuity and STRING, have specified the BMP antagonists as molecular hubs of interest in our dataset (2nd layer of analysis; Fig. 3.3.). Therefore, these bioinformatic tools have been utilized in a complementary fashion and successfully narrowed the high candidate number down to the BMP-antagonist signature, as potential marker of cancer progression in CRC desmoplastic CIFs.

It is generally strongly advocated that large datasets, either these are transcriptome- or proteome-based should be organized through more than one independent pathway analysis algorithm, coupled to user-friendly visualization tools, to ensure data readability and reproducibility, enhance network robustness, and reduce database- and operator-specific biases.

3.4.3. The Signature of Bone Morphogenetic Protein Inhibitors in DPD

In this study, we identified BMP antagonists/inhibitors as molecular hubs of the TGF-beta superfamily cluster, in form of a collective signature. As investigated further, the BMPIs identified here were either
highly-selective antagonists against certain BMP ligands (for instance, GREM1 covalently binds to and heterodimerizes with BMP2, BMP4 and BMP7)\textsuperscript{301}, or less-selective antagonists (for instance, FST and FSTL3 bind to BMPs with less affinity)\textsuperscript{507}. Moreover, other identified BMP inhibitors could bind to and modify extracellular components through proteolytic events, thus sequestering certain BMPs and reducing their bioavailability for receptor binding and complex formation (for instance, HTRA3 can induce decorin- and MGA-mediated BMP sequestration)\textsuperscript{318}. Interestingly, despite the reported differences, they all clustered up in our pathway analyses in one single protein-protein interaction network. Given the experimental evidence that BMP antagonists/inhibitors may be exerting coordinated activity\textsuperscript{276}, we faced the possibility to investigate them as signature, rather than individual regulators of the BMP pathway.

The BMP pathway, a developmental pathway which promotes the sustenance of epithelial phenotype in various cell types and cell lineages, may modulate tumor-suppressive effects in CRC and other cancers\textsuperscript{343, 355, 508-510}. Recent data suggest that sporadic CRC may disrupt the tumor-suppressive effects of BMP signaling by genomic mutations of SMAD4 and BMP receptor type II (BMPR2)\textsuperscript{359}. Other lines of evidence suggest the deployment of epigenetic mechanisms for the disruption of BMP signaling in tumors of the gastrointestinal (GI) tract\textsuperscript{511}. In the current study, as well as in others\textsuperscript{309}, the paracrine production of BMP antagonists, especially GREM1, served as a documented mechanism for BMP2/7 impairment and promotion of tumor cell motility and proliferation. Reasonably, certain questions could be raised from all these observations. The most important though is why do certain types of cancer, such as CRC, tend to develop a wide spectrum of mechanisms to block the tumor-suppressive properties of BMP signaling? To partially answer this question, we speculate that the tumor microenvironment in CRC deploys a plethora of strategies, most of the times serving as “safety valves”, to ensure that BMP signaling will be totally blocked. Here, we refer to the totality of these strategies as “the multilayered BMP-inhibition hypothesis”. This working hypothesis states that CRC cells of the advanced invasion front can only flourish in a microenvironment devoid of the tumor-suppressive properties of the BMP signaling. To guarantee its absolute disruption, the cancer cells deploy multiple “layers” of BMP antagonism, including the genetic level (i.e. SMAD4/BMPR2 functional mutations), the epigenetic level (i.e. methylation of BMP2 gene promoter), as well as the ligand level (i.e. antagonistic regulation of BMP ligands by paracrine production of BMP antagonists)\textsuperscript{345}.

A proof-of-concept observation that validates this working hypothesis comes from recent studies on the implication of receptor tyrosine kinase-like orphan receptor 2 (ROR2) in CRC. It appears, that there is documented lack of decoy-receptor-based BMP inhibition in CRC, which eventually allows high bioavailability of idle BMP receptors to dimerize with each other\textsuperscript{512}. Indeed, ROR2 is very frequently downregulated in CRC by aberrant promoter hypermethylation\textsuperscript{512}. Despite this epigenetic silencing could
have a paradoxical impact by assisting the perseverance of an active BMP signaling module, it appears that ROR2 downregulation facilitates non-canonical Wnt-mediated tumor promotion, which is especially important for CRC initiation and progression$^{512,513}$. As such, the CRC microenvironment could potentially deploy the multilayered-BMP inhibition signature, to overcome and bypass the paradoxical counter-effects rising from the absence of ROR2-mediated BMP suppression.

3.4.4. Conclusive Remarks

In the second chapter of this thesis, we have applied a two-layered pathway analysis strategy, based on modern bioinformatics approaches, in particular coupling together molecular concept and protein-protein interaction analyses, with purpose of delineating candidate mediators of local invasion and tumor-budding formation. Our analyses led us to the identification and characterization of a secreted signature, consisting of selective antagonists and/or other ECM-regulating inhibitors of the BMP pathway, as potential inducers of dedifferentiation and migration at the level of the desmoplastic CIF. For future perspective, we propose to establish expressional and mechanistic association between these BMPIs and cancer progression and metastasis, and possibly suggest novel targets for therapeutic interventions.
CHAPTER 4

Validating the Overrepresentation of a Secreted Signature of Bone Morphogenetic Protein Inhibitors in the Colorectal Cancer Desmoplastic Invasion Front

Sections of this chapter are pending publication:


Copyright permission has been granted.
4.1. INTRODUCTION

Previously we have developed *in vitro* cocultures between colon cancer cells and cancer-associated fibroblasts, in an attempt to mimic the paracrine interactions observed in desmoplastic boundaries, and further performed proteomic analysis in the obtained supernatants to decipher a proteomic interface of the reciprocal interactions, with ultimate aim of identifying potential regulators of local invasion and tumor budding formation \(^{453}\). The generated dataset, high-throughput in its very nature, was subsequently subjected to bioinformatic prioritization, based on the simple principle of pathway analysis, and a group of bone morphogenetic protein antagonists/inhibitors was opted as an attractive candidate for mediating these tumorigenic phenomena at the desmoplastic invasion front \(^{514}\).

Our next reasonable step will be to investigate the expressional and mechanistic associations of the proposed bone morphogenetic protein antagonists/inhibitors and establish links of correlation and causation with dedifferentiation and migration observed in the desmoplastic invasion front. However, it is generally known that the current status of high-throughput technologies, such as mass spectrometry-based proteomics and bioinformatics lack the appropriate sensitivity and specificity that would render them as highly-reproducible and concrete experimental platforms \(^{166}\). The future might be bright for this type of technologies, but currently, additional verification/validation experiments are usually essential to provide accuracy in the results originating from relevant pipelines \(^{515,516}\).

Overall, this chapter involves the exploitation of verification experiments using traditional molecular biology techniques and tissue-wide expression studies that together bridge and complement our proteomic and bioinformatic investigations.
4.2. EXPERIMENTAL PROCEDURES

4.2.1. Retrieval of Proteomic and Transcriptomic Datasets

Label-free quantification was performed on selected proteins (i.e. BMPIs) identified in the desmoplastic protein dataset (DPD), which is the proteomic dataset derived from secretome analysis of cocultures of SW480/SW620 cells with normal colonic fibroblasts 18Co. The dataset is described on chapter 2, and all the corresponding proteins are listed on Table A (Appendix; Table A). Gene expression data were retrieved from Navab et al. (2011) 499, Omnibus (GEO) database, accession no. GSE22874.

4.2.2. Real-Time PCR

The human SW480, SW620 and 18Co cell lines were purchased from the American Type Culture Collection (Rockville, MD) and were maintained in DMEM with 10% FBS. Cocultures of the “indirect” type were generated for investigation of BMPI expression. In brief, we stimulated SW480 and SW620 cancer cells with 18Co fibroblast medium and vice-versa. Stimulation media (SW480, SW620 and 18Co) were collected under serum-free conditions, as previously described 127. Stimulated cells were kept for 2 days under media containing 50% stimulation medium and 50% fresh medium. Following the termination of all experiments, cells were collected using non-enzymatic dissociation buffer (EDTA), and subjected to RNA extraction and cDNA synthesis, as described 184, 517, 518. RT-PCR was performed by using 1X SYBR green reagent (Applied Biosystems) and transcript levels of GREM1, FST and HTRA3 were measured on a 7500 ABI System. The following 5’-3’ forward (F) and reverse (R) primer sequences were used: 
GREM1- F: 5’-TCATCAACCGCTTCTGTTACG-3’, R: 5’-CAGAAGGAGCAGGACTGAAAGG-3’
FST- F: 5’-AGCAGCCAGAACTGGAAGTCCA-3’, R: 5’-CCGATTACAGGTCACACAGT-3’
HTRA3- F: 5’-CAAGAAGTGGGACATTGCCACC-3’, R: 5’-CTTTGTCACTGTTTCTGT-3’
GAPDH- F: 5’-GTCTCCTCTGACCTCAACACGG-3’, R: 5’-ACCACCTGTGTTCAGGCCA-3’

4.2.3. GREM1 Immunoassay

SW480, SW620 and 18Co cells were seeded in triplicates in 6-well plates (DMEM, 10% FBS), either in monocultures or direct cocultures (SW480Co/SW620Co) up to ~50-60% confluence, as described elsewhere 357. CM from all wells were then replaced with 2 ml serum-free DMEM. CM were collected
after 2 days and GREM1 concentration was measured with a commercially available immunoassay (Invitrogen), using the manufacturer’s instructions.

4.2.4. Immunohistochemistry

A patient cohort of 30 invasive colon cancers, encompassing well-, moderately-, and poorly-differentiated adenocarcinomas (n= 3, 16 and 11, respectively) was kindly provided by Dr. Robert H. Riddell and Dr. Richard Kirsch and was used in our previous study. Staining was carried out with avidin-biotinylated peroxidase complex (ABC Elite Kit, Vector Laboratories) on a LabVision 720 Autostainer (Labvision, Fremont, CA) using 1X TBS-Tween rinse and wash solution. Four μm thick paraffin sections were mounted on positively charged glass slides and baked at 60°C for 60 min. Sections were then immersed in xylene, followed by graded alcohols, at 10 min per solution. Antigen retrieval involved heating slides in a citrate (pH 6.0) buffer in either a microwave or a pressure cooker. Sections were then incubated with the primary antibodies for gremlin-1 (GREM1; rabbit polyclonal; dilution 1:200; Abcam), follistatin (FST; rabbit polyclonal; dilution 1:50; Sigma), high-temperature requirement-A3 (HTRA3; rabbit polyclonal; dilution 1:300; Sigma), alpha-smooth muscle actin (alpha-SMA; Clone 1A4, dilution 1:200, Dako), h-caldesmon (h-CALD; dilution 1:400; Dako), laminin-beta1, (LAMB1; rabbit polyclonal; dilution 1:400; Sigma), and beta-catenin (beta-CAT; clone 14; 1:200; BD Bioscience). Diaminobenzidine hydrochloride solution with hydrogen peroxide (Sigma) was the chromogen. Briefly, the beta-catenin slide from each patient was used for selecting the areas in an unbiased fashion, to be eventually scored with the BMPIs. Slides corresponding to alpha-SMA, LAMB1 and h-CALD were then used to collectively demonstrate desmoplastic or non-desmoplastic nature in the peritumoral stroma of the selected areas. The exact protocols and scoring system, by which BMPI expression was evaluated, as well as the method by which beta-catenin and the three CAF markers were used are detailed in chapter 5.

4.2.5. Statistical Analysis

The SPSS (version 20) software was used for all analyses. Data are presented either as raw values (i.e. spectral counts of individual proteins) or as bars of mean values with standard errors (i.e. ELISA, IHC data), or as relative values (i.e. ratios) between two experimental conditions (i.e. RT-PCR in indirect cocultures). Comparison between mean/median values of two matched/paired samples (i.e. IHC data) was performed with the non-parametric Wilcoxon statistic. Significance was shown at the 0.05 level.
4.3. RESULTS

4.3.1. In Vitro Overexpression of Bone Morphogenetic Protein Inhibitors, as a Result of Cancer Cell-Cancer-Associated Fibroblast Interactions

On the basis of identifying a small group of BMPIs, as promising regulators of local invasion and tumor budding formation, here, we provided some in vitro validation of our original proteomic experimentation. Indeed, the normalized spectral counts of FST, FSTL3, GREM1 and HTRA3, the major molecular hubs that were overrepresented in the TGF-beta superfamily cluster, revealed that FSTL3 and GREM1 were found in both cocultures, whereas FST and HTRA3 were identified in the SW620Co coculture only, while none of these four BMPIs were identified in the monoculture controls (Fig. 4.1A.).

Based on the notion that MS-related proteomic data lack adequate sensitivity and specificity, despite the major technological accomplishments in the modern oncoproteomics arena, we first reproduced the same mono- and cocultures of the original experiment and measured GREM1 levels with a commercially-available immunoassay in 2-day generated CM. Indeed, GREM1 was secreted by both SW480Co and SW620Co cocultures (~15 ng/mL), while only basal levels were observed in the monocultures, verifying to some extent the MS-related data for this protein (Fig. 4.1B.).

Moreover, when we stimulated 18Co cells with 2-day supernatants derived from either SW480 or SW620 cell lines, we observed ~4-fold increase in GREM1, ~2.8-fold increase in FST, and a 3-4-fold increase in HTRA3 gene expression levels (Fig. 4.3C.). The reverse scenario, i.e. the stimulation of SW480 and SW620 colon cancer cells with 2-day supernatant derived from 18Co cells resulted in very small increases of <1.3-fold in all cases (Fig. 4.3C.). These observations collectively suggest that colonic fibroblasts may possibly increase BMPI gene expression in a paracrine fashion.

As a proof-of-concept, we utilized publicly-available microarray data from a biological context. In this study, transcriptome profiles of cell lines derived from either CAFs or NFs of patients with non-small lung cell carcinoma (NSCLC), were compared. We looked into the supplementary data of this experiment, and found that ~6.4-fold upregulation of GREM1 in CAFs when compared to NFs was noted (Fig. 4.3C.). Thus, despite seen in a different type of cancer (i.e. lung cancer), it seems that the desmoplastic microenvironment is principally responsible for GREM1 secretion in the tumor-host cell interface. No data for FST and HTRA3 were made available in this dataset.
FIGURE 4.1. In vitro overexpression of BMP inhibitors, as a result of tumor host-cell interactions. (A) Total spectral counts of GREM1, FST, FSTL3 and HTRA3 in three monoculture supernatants (SW480, SW620, 18Co) and the respective coculture supernatants (SW480/18Co and SW620/18Co). Proteomic data for this analysis were derived from our previous study. (B) Gremlin-1 protein levels were assessed through ELISA in conditioned media, after recapitulating the experimental conditions described in (A). Note that only basal levels of GREM1 were detected in monoculture supernatants using a specific GREM1 ELISA. (C) Relative GREM1, FST and HTRA3 gene expression levels in indirect cocultures (stimulation with conditioned media). The last bar in the graph depicts differences in GREM1 expression levels between cancer associated fibroblasts and normal fibroblasts in the Navab study. Bars demonstrate fold-differences of GREM1, FST and HTRA3 gene expression in stimulated versus non-stimulated cells.
4.3.2. Bone Morphogenetic Protein Inhibitors GREM1 and HTRA3 Are More Highly Expressed in Desmoplastic, Compared to Normal Stromata in Colorectal Cancer Patients

Besides verifying *in vitro*, the BMPI overexpression after reproducing the same or relevant experimental setups as the original proteomic platform, we also provided pathophysiological relevance of BMPI expression in actual CRC patients, by using human pathology (immunohistochemical) approach. To achieve that, we retrieved a patient cohort of 30 invasive CRC cases, which had been used for evaluating the expression of multiple tight junction proteins, in our previous study 127. Of these, a total of 21 cases had an adequate representation of adjacent healthy tissues, involving normal colonic mucosa with underlying submucosal and muscular layers (Fig. 4.2A.). To demonstrate whether these BMPIs are overrepresented in the desmoplastic interface, we compared GREM1, FST and HTRA3 expression levels between pathologic/desmoplastic stromata and adjacent normal control stromata from the same cases (Fig. 4.2A.). Desmoplastic stromata were identified by investigating the adjacent-to-the-tumor, stromal areas for the well-established myofibroblastic signature (alpha-SMA positive; LAMB1 positive; h-CALD negative), and any stromata that did not relate to the above-mentioned signature were excluded from further analysis. Our analysis demonstrated significant (p<0.05; Wilcoxon test) increase of GREM1 and HTRA3 expression in 21 desmoplastic (GREM1, Mean=2.32, SD=0.38; FST, Mean=1.38, SD=0.32; HTRA3, Mean=2.17, SD=0.39) stromata compared to 21 matched normal (GREM1, Mean=0.23, SD=0.38; FST, Mean=1.13, SD=0.35; HTRA3, Mean=1.06, SD=0.32) stromata (Fig. 4.2B.). Representative snapshots of one patient, showing increased GREM1 expression in desmoplastic lesions around the tumor, as opposed to matched normal stroma is provided (Fig. 4.2C.).
FIGURE 4.2. Verification studies in a CRC patient cohort, using immunohistochemistry. (A) Schematogram and immunohistochemical picture, explaining the design of the verification experiment. Areas from which normal and pathologic stromata were selected for further scoring are shown squares in either the cartoon-style illustration (left panel) or the beta-catenin stain IHC figure (right panel). (B) Mean expression scores of GREM1, FST and HTRA3 in pathologic and normal stromata of 21 CRC patients. Asterisks denote statistical significance, p<0.05, Wilcoxon test. (C) Representative IHC snapshots from one patient, demonstrating evident difference in the stromal expression of GREM1 between normal and pathologic stromata. Note the slight GREM1 expression around the colonic crypts of adjacent healthy mucosa (positive control in the picture on the left). All magnifications X200; except for last X400.
4.4. DISCUSSION

4.4.1. Importance of Incorporating Verification Experiments in OMICS-associated Platforms

Secretome analysis has only recently been established as a sub-field of oncoproteomics and the indications thus far point to the fact that this source of proteins is a promising pool of biomarkers and therapeutic targets for various types of cancer. There is a reasonable consensus that efforts should continue to comprehensively analyze the cancer secretome. Secreted proteins account for approximately 10-15% of the proteins encoded by the human genome and participate in various physiological processes such as immune defense, blood coagulation, matrix remodeling and cell signaling, but also in pathological conditions including cancer angiogenesis, differentiation, invasion and metastasis \(^\text{166}\).

It is true that despite major technological advances in the field, secretome analysis using mass spectrometry has not yet managed to obtain the experimental robustness and accuracy of more concrete and established platforms, such as for instance, gene expression analysis using microarrays \(^\text{165, 519}\). However, not all biological questions can be addressed in the level of genome and transcriptome, and as such, scientific evidence should definitely be based on protein analysis methods for the characterization and quantification of the proteome, which may be of particular importance in certain diseases, like cancer \(^\text{445, 519}\). Over time, it has become quite archetypal that proteomic studies, which are based on bottom-up or top-down approaches for protein identification (i.e. in cell lines, tissues and fluids), should always include a verification step to avoid false-positive results. As such, experiments presented in this chapter aimed to provide some verification potential, (I) by demonstrating that our results could be reproduced in \textit{in vitro} cocultures, and also (II) by incorporating a pilot study, using immunohistochemistry.

4.4.2. Conclusive Remarks

In this chapter, we have provided both gene- and protein-expression evidence that the identified and prioritized BMPIs from our previous observations (i.e. chapters 2 & 3), especially GREM1, are induced through tumor-host cell interactions, in particular the paracrine interactions between cancer cells and cancer-associated fibroblasts. The remainder of this thesis will explicitly focus on the contextual and mechanistic associations of GREM1/HTRA3 with local invasion, tumor-budding, and epithelial-to-mesenchymal transition at the CRC invasion front.
CHAPTER 5

Expression Pattern of Bone Morphogenetic Protein Inhibitors in the Colorectal Cancer Desmoplastic Invasion Front

Sections of this chapter have been published in Molecular Oncology:


Copyright permission has been granted.
5.1. INTRODUCTION

Our previous investigations in an in vitro coculture model system, recapitulating the paracrine interactions between colon cancer cells and cancer-associated fibroblasts, have pointed out the overrepresentation of multiple members of the BMP/TGF-beta signaling pathways, including inhibitory and/or antagonist proteins, like, for example, gremlin-1, follistatin, follistatin-like protein-3, high temperature requirement-A3 (or probable serine protease-A3) and fibromodulin 453,514.

Given that such paracrine interactions in desmoplastic interfaces are known to mediate local invasion and tumor-budding formation by promoting cell biological programs that provide migratory/invasive opportunities to cancer cells (i.e. epithelial-to-mesenchymal transition) 110,154, we further hypothesized that this secreted signature of BMP antagonists/inhibitors, here referred to as “BMPI”, could be involved in such molecular events.

To provide insights into this working hypothesis, we first decided to examine the contextual correlation of BMPIs with the desmoplastic invasion front. As such, this chapter describes the design and undertaking of an immunohistochemical approach to globally describe the expression profiles of three BMPIs, in particular GREM1, FST and HTRA3, in two independent (discovery and validation) patient cohorts of invasive colorectal cancer.
5.2. EXPERIMENTAL PROCEDURES

5.2.1. Patient Cohorts

The study involved two distinct patient cohorts, both randomly obtained from a larger well-established cohort of 219 patients with invasive CRC, first described in our previous study \(^{520}\). The discovery cohort (n=30) was blindly selected with respect to creating three independent groups based on histological criteria defining the tumor-budding population, as previously described \(^{153}\). In particular, high-grade tumor-budding (n=10), low-grade tumor-budding (n=10) and tumor-budding-free (n=10) patient groups were selected. The validation cohort (n=38) had the following clinicopathologic characteristics: (1) median follow-up of survivors: 56 months (range 0-79; interquartile range 14-64); (2) median age of cohort: 67.5 years (range 34-89; interquartile range 60-77); (3) percentage of males: 55.3% (n=21); (4) disease location: of the 20 right sided cancers (52.6%), 11 (28.9%) were located in the caecum, 6 (15.8%) in the ascending colon and 2 (5.3%) in the transverse colon and 1 (2.6%) in the splenic flexure. Of the 18 left-sided cancers (47.4%), 2 (5.3%) were located in the descending colon, 12 (31.6%) in the sigmoid colon and 4 (10.6%) at the recto-sigmoid junction/upper rectum; (5) 30-day mortality: 5.3% (n=2); (6) T-Category: 30 (78.9%) were classified as pT3 and 8 (21.1%) as pT4; (7) the 5-year disease-free survival (DFS) was 68.1% and the 5-year overall survival (OS) was 81.6%; (8) Regarding tumor-budding, 14 cases (36.8%) were identified with medium/high tumor-budding (TB > or = 5); (9) Crohn’s like reaction was present in 5 cases (13.1%); (10) marked tumor-infiltrating lymphocytes (marked-TILs) were found in 7 cases (18.5%).

5.2.2. IHC Staining

Staining was carried out, as described in chapter 4.2.5. The following primary antibodies were used: alpha-smooth muscle actin (alpha-SMA; Clone 1A4, dilution 1:200, Dako), h-caldesmon (h-CALD; dilution 1:400; Dako), laminin-beta1, (LAMB1; rabbit polyclonal; dilution 1:400; Sigma), beta-catenin (beta-CAT; clone 14; 1:200; BDBioscience), as well as the three tested BMP inhibitors GREM1 (rabbit polyclonal; dilution 1:200; Abcam), HTRA3 (rabbit polyclonal; dilution 1:300; Sigma) and FST (rabbit polyclonal; dilution 1:50; Sigma). Bone morphogenetic protein antagonists have been shown to be expressed by naturally-occurring myofibroblasts around healthy colonic crypts, whereby they support the stem-cell niche and the differentiation gradient of the normal colonic epithelium along the crypt axis \(^{64,521}\). As such, the pericryptal expression of GREM1, HTRA3 and FST served as positive internal controls. Optimization experiments ensured minimal background staining and silenced cells/tissues (i.e. smooth
muscle cells) served as negative controls in each sample independently (Appendix; Fig. C). Additional negative controls were performed by omitting the primary antibody step.

5.2.3. Selection of Immunohistochemical Areas (i.e. “Contexts”) for Scoring

Based on the concept of “tumor development along with the surrounding context”, described by Mina Bissell group 4, as well as our previous work on CRC desmoplasia 453, here, we assumed that the expression of BMP antagonists GREM1, HTRA3 and FST is restricted in specialized microenvironments, such as the desmoplastic invasion front, whereby features of localized invasion such as tumor budding formation are mediated. Therefore, a small number (n=1-6) of sections in each specimen, encompassing such microenvironments, here referred to as “contexts”, was randomly selected in each case (Fig. 5.1.), based on beta-CAT staining, which provides adequate discriminatory power between cancer and stromal cell populations (Fig. 5.2A.). Contexts encompassing tissue lifting, edge artifacts, high background staining and abundant mucinous component were excluded. The desmoplastic nature of the acquired contexts was verified with the myofibroblastic markers alpha-smooth muscle actin (alpha-SMA) and laminin-beta-1 (LAMB1), as well as the smooth-muscle marker h-caldesmon (h-CALD). Only peritumoral stromata, bearing an expression signature reminiscent of the desmoplastic microenvironment (i.e. alpha-SMA positive, LAMB1 positive, h-CALD negative or focally-positive) was accepted for further consideration in this study (Fig.5.2B.) A total of 90 and 167 contexts fulfilled these quality control criteria for the discovery and validation cohorts, respectively.
FIGURE 5.1. Study design and context selection. Schematic illustration, demonstrating tumor invasion fronts with (red squares) or without (yellow squares) tumor budding populations. Each one of these selected contexts (see figure 5.2. for context selection and characterization) was independently scored for epithelial/stromal expression of BMPs.
FIGURE 5.2. **Context selection and characterization.** Optimization experiments with immunohistochemical markers for defining tumor-budding populations and desmoplastic lesions. (A) **Discriminatory power of beta-catenin staining in desmoplastic invasion fronts.** Note the ability of beta-CAT in discriminating between cancer and stromal populations. As such, beta-CAT is efficient in the identification of tumor budding areas contexts (e.g. black arrow) in desmoplastic lesions. All magnifications x100. (B) **Defining of CAF population and peritumoral desmoplastic lesions.** Confirmation of desmoplastic stroma was performed through the combination of three individual cell-based markers, namely alpha-SMA, h-CALD and LAMB1. A figure showing early invasion was selected to demonstrate briefly the relative expression of these cell-based markers in various subpopulations that can be most commonly encountered in the microenvironment of CRC. The panel below demonstrates the signature in each cell type of interest. All magnifications x200.

<table>
<thead>
<tr>
<th>Code</th>
<th>Cell-Type</th>
<th>alpha-SMA</th>
<th>h-CALD</th>
<th>LAMB1</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>normal fibroblasts</td>
<td>-</td>
<td>-</td>
<td>+/-</td>
</tr>
<tr>
<td>b</td>
<td>normal pericryptal myofibroblasts</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>normal colonocytes</td>
<td>-</td>
<td>-</td>
<td>+/-</td>
</tr>
<tr>
<td>d</td>
<td>cancer cells</td>
<td>-</td>
<td>-</td>
<td>+/-</td>
</tr>
<tr>
<td>e</td>
<td>immune/inflammatory cells</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>f</td>
<td>smooth muscle cells (muscularis mucosae)</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>g</td>
<td>cancer-associated fibroblasts</td>
<td>+</td>
<td>-</td>
<td>+/-</td>
</tr>
</tbody>
</table>

* A small subgroup of h-CALD+ CAFs has been reported. Thus, some focal staining of h-CALD in desmoplastic lesions does not exclude the presence of CAFs.
5.2.4. Scoring of BMPIs

Of the 90 contexts of the discovery cohort, a total of 35 had tumor-budding populations and were named: “tumor-budding areas”, while 55 did not, and were named: “invasion fronts” (Table 5.1.). Using histological criteria described elsewhere, the contexts were classified according to their tumor-budding score (TBS) into: “no tumor-budding” (NTB; 0<TBS<5; n=55), “low-grade tumor-budding” (LGTB; 5.1<TBS<10; n=18) and “high-grade tumor-budding” (HGTB; 10.1<TBS<15; n=17) (Table 5.1.).

GREM1, HTRA3 and FST were scored for both epithelial (e) and stromal (s) compartment of the tumors in all 90 contexts, either these simply represented invasion fronts or tumor-budding areas. Therefore, the following scores were obtained: eGREM1, sGREM1, eHTRA3, sHTRA3, eFST and sFST, here described: “BMPI determinants”. To obtain a single numerical value for each of these determinants, two independent parameters were co-assessed: the staining intensity (SI) and the staining percentage (SP). For the BMPI determinant under investigation, the SI was scored with 0 (no intensity), 1 (low intensity), and 2 (strong intensity) (Fig. 5.3.). Accordingly, the SP was scored with 0 (0% of area is positive), 1 (0.1%-10% of area is positive), 2 (10.1%-50% of area is positive), and 3 (50.1%-100% of area is positive). The two parameters were multiplied to provide the final score for each BMPI determinant in each context [e.g. eGREM1=(SI)eGREM1*(SP)eGREM1]. From the range of the numerical values of the SI and SP parameters, it follows that each determinant could be designated with a score of 0-6.

Total scores for GREM1 (tGREM1), HTRA3 (tHTRA3) and FST (tFST), were calculated by combining the individual scores for (e) and (s) BMPI determinants [e.g. tGREM1=eGREM1+sGREM1]. Thus, each of the tGREM1, tHTRA3 and tFST values could be designated with a score of 0-12. Finally, for each context, the total BMPI (tBMPI) index was calculated from scores of all three BMPIs [tBMPI=tGREM1+tHTRA3+tFST]. Thus, tBMPI obtained a numerical value of 0-36 in each context. Based on the tBMPI scores, we classified the contexts into Q1 (n=22; tBMPI=0-3; low BMPI-expressing quartile), Q2 (n=26; tBMPI=4-7; low-medium BMPI-expressing quartile), Q3 (n=21; tBMPI=8-11; medium-high BMPI-expressing quartile) and Q4 (n=21; tBMPI=11-36; high BMPI-expressing quartile). Each one of these quartiles represents ~25% of the 90 contexts in total (Table 5.1.). It follows that this classification may categorize the contexts into varying levels/degrees of BMP antagonism, regardless of which individual BMPIs contribute to the overall BMP antagonism.

Of the 167 contexts of the validation cohort, a total of 89 had tumor budding, while the remaining 78 were invasion fronts, free of tumor-budding populations. This cohort was used for establishing the difference of eHTRA3 expression between these two types of contexts. The scoring of eHTRA3 was performed as described above.
<table>
<thead>
<tr>
<th>Nature of contexts</th>
<th>Frequency (N)</th>
<th>Percent (%)</th>
<th>Cum. Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invasive fronts</td>
<td>55</td>
<td>61.1</td>
<td>61.1</td>
</tr>
<tr>
<td>Tumor-budding areas</td>
<td>35</td>
<td>38.9</td>
<td>100.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tumor-budding grade</th>
<th>Frequency (N)</th>
<th>Percent (%)</th>
<th>Cum. Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTB</td>
<td>55</td>
<td>61.1</td>
<td>61.1</td>
</tr>
<tr>
<td>LGTB</td>
<td>18</td>
<td>20.0</td>
<td>81.1</td>
</tr>
<tr>
<td>HGTB</td>
<td>17</td>
<td>18.9</td>
<td>100.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>tBMPI expression score quartiles</th>
<th>Frequency (N)</th>
<th>Percent (%)</th>
<th>Cum. Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 (tBMPI=0-3; Mean=1.73)</td>
<td>22</td>
<td>24.4</td>
<td>24.4</td>
</tr>
<tr>
<td>Q2 (tBMPI=4-7; Mean=5.65)</td>
<td>26</td>
<td>28.8</td>
<td>53.2</td>
</tr>
<tr>
<td>Q3 (tBMPI=8-11; Mean=9.24)</td>
<td>21</td>
<td>23.4</td>
<td>76.6</td>
</tr>
<tr>
<td>Q4 (tBMPI=11-26; Mean=16.95)</td>
<td>21</td>
<td>23.4</td>
<td>100.0</td>
</tr>
</tbody>
</table>

*Type of context, as defined by beta-CAT staining
**Grade defined based on the histological criteria described Mitrovic et al. (2012)
***Abbreviations: NTB, no tumor-budding; LGTB, low-grade tumor-budding; HGTB, high-grade tumor-budding
****Classification of contexts into quartiles based on the context’s tBMPI expression score
FIGURE 5.3. BMP inhibitor expression scoring system. For simplifying the study design, a three-level scoring system (ranging from 0 to 2) was applied for measuring immunoreactivity of all BMPIs (GREM1 is shown here) in the desmoplastic invasion fronts with or without tumor budding. Immunoreactivity was scored as relative intensity to the one shown in the figures above. Staining intensity was independently assessed in the stromal and epithelial compartments in each context. All magnifications x200.
5.2.5. Statistical Analysis

The SPSS (version 20) software was used for all analyses. Data on expression levels of various BMPIs were presented as means with their standard error.

The Kolmogorov-Smirnov statistic was performed to test for normality of distribution in the BMPI score values. In cases where data did not follow a normal distribution, the non-parametric Mann-Whitney U-test was utilized to compare BMPI scores between two independent groups of contexts. In cases where data followed a normal distribution, the parametric independent-samples student’s t-test was utilized instead. In these cases, the Levene’s test was a priori performed to test for equality of variances between the two groups. Whenever equal variances could not be assumed, the Levene’s correction in degrees of freedom was performed for correcting the obtained p-values. The non-parametric Wilcoxon test was used to compare BMPI scores between two matched/paired groups of contexts. The non-parametric Jonckheere-Terpstra test was used for comparing individual BMPI scores among multiple independent groups of contexts. For correlations, the non-parametric Spearman’s ranked correlation coefficient was applied. Correlation matrices for multiple variables were created using Spearman’s rho values. Statistical significance was shown at either 0.05 or 0.01 level.

For logistic regression analysis, prediction criterion was a dichotomous variable indicating the presence of contexts in CRC invasion fronts, either designated as “tumor-budding areas” (N=35) or “invasion fronts” (N=55). Factor included in the analysis was the epithelial expression of the BMP antagonist HTRA3 (i.e. eHTRA3). A p-value<0.05 was regarded as statistically significant. Goodness-of-fit was a priori performed with the Hosmer-Lemeshow (HL) statistic; p-value<0.05 was considered significant to show improper data calibration. Receiver operating characteristic (ROC) curves were used to assess the adequacy of the predictive power. Cross-tabulation was used to assess sensitivity and specificity. For validation purposes, the logistic regression model was performed using eHTRA3 expression scores from the validation cohort. In the latter case, a series of co-factors such as age, gender, T-category, Lymph node yield, status of intravascular invasion, anatomical site, presence of Crohn’s like lesions and marked tumor-infiltrating lymphocytes (marked-TILs) were additionally included in the classification model.

The standard deviation of eHTRA3 was calculated in the discovery cohort to assist in the determination of sample size for the validation cohort. A detectable difference of eHTRA3 expression equal to 1 standard deviation at the 0.05 significance level was requested for our study. Using this criterion, the Harvard sample size calculator (http://hedwig.mgh.harvard.edu/sample_size/size.html) demonstrated that we need 36 patients, in order to achieve 90% (i.e. 0.9) power in our study. We, thus, incorporated 38 individuals to ensure the probability of making type II error was <10%.
5.3. RESULTS

5.3.1. BMPI Expression in Colorectal Cancer Desmoplastic Invasion Fronts

In each of the 90 selected contexts in the discovery cohort, six independent scores corresponding to either stromal (s) or epithelial (e) expression of GREM1, HTRA3 and FST, were provided. Descriptive statistics for eGREM1, sGREM1, eHTRA3, sHTRA3, eFST and sFST, here termed as “BMPI determinants”, are shown in Table 5.2. Interestingly, all BMPI determinants were expressed in both epithelial and stromal compartments of CRC desmoplastic invasion front. The highest mean expression score was noticed for HTRA3 (tHTRA3; mean, 3.37; standard error, 0.27), followed by GREM1 (tGREM1; mean, 2.46; standard error, 0.284) and FST (tFST; mean, 2.35; standard error, 0.193). Total BMPI values (tBMPI; mean, 8.17; standard error, 0.626) (Table 5.2.) defined 4 quartiles (Q1-Q4) of contexts, corresponding to increasing levels of BMP antagonism (Table 5.2.). In a first attempt to characterize the BMPI expression signature in CRC, we investigated the expression pattern of six BMPI determinants across increasing levels of BMP antagonism. The signature tended to retain a similar pattern of BMPI expression from low-to high-BMPI-expressing quartiles of contexts (dashed red lines demonstrate the mean tBMPI in each quartile) (Appendix; Fig. D).
Table 5.2. Descriptive statistics from scoring of three BMPIs in the discovery cohort (N=90)

<table>
<thead>
<tr>
<th>BMPIs</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Error</th>
<th>Std. Dev</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>eGREM1</td>
<td>0</td>
<td>6</td>
<td>1.61</td>
<td>0.18</td>
<td>1.707</td>
<td>2.914</td>
</tr>
<tr>
<td>sGREM1</td>
<td>0</td>
<td>6</td>
<td>0.84</td>
<td>0.138</td>
<td>1.306</td>
<td>1.706</td>
</tr>
<tr>
<td>tGREM1 **</td>
<td>0</td>
<td>10</td>
<td>2.46</td>
<td>0.284</td>
<td>2.691</td>
<td>7.24</td>
</tr>
<tr>
<td>eHTRA3</td>
<td>0</td>
<td>6</td>
<td>1.01</td>
<td>0.16</td>
<td>1.518</td>
<td>2.303</td>
</tr>
<tr>
<td>sHTRA3</td>
<td>0</td>
<td>6</td>
<td>2.36</td>
<td>0.162</td>
<td>1.538</td>
<td>2.367</td>
</tr>
<tr>
<td>tHTRA3 **</td>
<td>0</td>
<td>12</td>
<td>3.37</td>
<td>0.27</td>
<td>2.564</td>
<td>6.572</td>
</tr>
<tr>
<td>eFST</td>
<td>0</td>
<td>6</td>
<td>1.77</td>
<td>0.153</td>
<td>1.454</td>
<td>2.113</td>
</tr>
<tr>
<td>sFST</td>
<td>0</td>
<td>3</td>
<td>0.58</td>
<td>0.087</td>
<td>0.821</td>
<td>0.674</td>
</tr>
<tr>
<td>tFST **</td>
<td>0</td>
<td>7</td>
<td>2.35</td>
<td>0.193</td>
<td>1.831</td>
<td>3.352</td>
</tr>
<tr>
<td>tBMPI ***</td>
<td>0</td>
<td>26</td>
<td>8.17</td>
<td>0.626</td>
<td>5.935</td>
<td>35.219</td>
</tr>
</tbody>
</table>

BMPI, bone morphogenetic protein inhibitor; GREM1, gremlin 1; HTRA3, high temperature requirement A3; FST, follistatin
** total expression was calculated by combining the s and e values of each individual BMPI (refer to: Materials & Methods)
*** total BMP antagonism was calculated by combining the t values of all BMPIs (refer to: Materials & Methods)
5.3.2. Coordinated Epithelial and Stromal Expression of BMPIs in Desmoplastic Invasion Fronts

First, we wished to determine the individual contribution of each BMPI determinant, as the overall levels of BMP antagonism in the desmoplastic invasion front increase. By comparing the expression of the six BMPI determinants among the Q1-Q4 contexts, we found statistically significant differences (p<0.01; Jonckheere-Terpstra test) in the ranked expression scores for all six of them (Fig. 5.4A). All BMPI determinants exhibited stable increase in their mean expression scores across the Q1-Q4 contexts (Fig. 5.4A.). Also, although Q1 contexts were almost deficient in eGREM1 expression, this BMPI determinant presented with the highest mean expression score among all investigated BMPI determinants in the contexts of high BMP antagonism (i.e. Q4 contexts) (Fig. 5.4A.).

Second, by correlating the total expression scores of the three BMPIs, we found statistically significant associations (p<0.01; Spearman’s ranked correlation coefficient) between all combinations of correlations (i.e. tGREM1/tHTRA3, tGREM1/tFST and tHTRA3/tFST) (Figure 5.4B.). In addition, when we performed correlation studies using the individual (s) and (e) scores, we again came across statistically significant associations (p<0.01 or p<0.05, refer to Fig. 5.4C. for details; Spearman’s ranked correlation coefficient) between all possible combinations of correlations (with the exception of certain correlations regarding the eHTRA3 determinant). As illustrated from the correlation matrices (Figs 5.4B. and 5.4C.), the calculated correlations were all positive (Spearman’s rho>0), signifying the coordinated increase of mean expression score values for all BMPIs.

Two representative contexts, the first (tGREM1=0; tHTRA3=2; tFST=0; tBMPI=2) derived from Q1 quartile, while the second (tGREM1=10; tHTRA3=7; tFST=5; tBMPI=22) derived from Q4 quartile are shown (Fig. 5.4D.) for proof-of-concept. Collectively, data presented in this section allowed us to conclude that, first, there is a pattern of BMPI co-regulation, and second, multiple BMPIs may be implicated in the overall BMP antagonism in CRC desmoplastic microenvironments, at the same time.
FIGURE 5.4. Coordinated epithelial and stromal expression of multiple BMPs in colorectal cancer desmoplastic invasion fronts. (A) Patterns of mean expression scores of BMP determinants across Q1-Q4 quartiles. Asterisks demonstrate statistically significant changes across the four groups with Jonckheere-Terpstra test (p<0.01) (B-C) Correlation matrices for all possible combinations of total expression scores of three BMPs (B) or individual (s) and (e) expression scores for all three BMPs (C). The numbers indicate Spearman’s rho for each combination. Asterisks demonstrate significance in either 0.05 level (**) or 0.01 level (*), with Spearman’s ranked correlation coefficient. (D) IHC snapshots from three selected contexts. The expression of BMPs with increasing levels of overall BMP antagonism is demonstrated. Magnifications x200.
5.3.3. Microenvironmental Polarity of BMPI Expression in Desmoplastic Invasion Fronts

The observed ‘peaks’ on the BMPI expression pattern across Q1-Q4 groups almost conservatively corresponded to three particular BMPI determinants, namely the eGREM1, sHTRA3 and eFST ones (Appendix; Fig. D; black arrows). This led us to hypothesize that the expression BMPIs participating in BMP antagonism is polarized, i.e. it is different but consistent for each BMPI on both sides of the tumor-host cell interface. The highest mean expression score in the stromal BMPI determinants was noticed for sHTRA3 (mean, 2.36; standard error, 0.162), followed by sGREM1 (mean, 0.84; standard error, 0.138) and sFST (mean, 0.58; standard error, 0.087) (Table 5.2.). Also, the highest mean expression score in the epithelial BMPI determinants was noticed for eFST (mean, 1.77; standard error, 0.153), followed by eGREM1 (mean, 1.61; standard error, 0.18) and eHTRA3 (mean, 1.01; standard error, 0.16) (Table 5.2.). Accordingly, both GREM1 and FST demonstrated a preferential expression in the epithelial compartments (p<0.05; Wilcoxon test), while HTRA3 in the respective stromal compartments of the tumors investigated (p<0.05; Wilcoxon test), regardless of context type (i.e. invasion front or tumor-budding area) (Figs. 5.5A-C.)

Microenvironmental polarity of GREM1/FST expression in the epithelial compartment was also demonstrated through significantly higher mean expression scores of both eGREM1 and eFST compared to eHTRA3 (p<0.05; Wilcoxon test), across increasing BMPI-expression score quartiles (Fig. 5.5D.). An exception to this observation was the Q4 group of contexts, whereby eHTRA3 almost reached the same expression levels with eGREM1 and eFST (p>0.05; Wilcoxon test) (Fig. 5.5D.). This occurred due to a significant increase of eHTRA3 mean expression score between Q3 and Q4 groups of contexts (p<0.05; Mann-Whitney U-test) (Fig. 5.5D.). On the other hand, sHTRA3 demonstrated a significantly higher mean expression score compared to either sGREM1 or sHTRA3 determinants (p<0.05; Wilcoxon test), across increasing BMPI-expression score quartiles, without any exceptions (Fig. 5.5E.).

A representative context (tGREM1=2; tHTRA3=2; tFST=6; tBMPI=10) was selected to illustrate the microenvironmental polarity of BMPI expression. Immunoreactivity of HTRA3 was solely detected in the stromal compartment (sHTRA3=2; eHTRA3=0), while immunoreactivity of GREM1 and FST was detected in the epithelial compartment (sGREM1=0; eGREM1=2; sFST=0; eFST=6) of the tumor (Fig. 5.5F.). These data briefly indicate that: (a) The major determinant of stromal BMP antagonism in CRC desmoplastic invasion fronts is HTRA3, and (b) The major determinants of epithelial BMP antagonism in CRC desmoplastic invasion fronts are GREM1 and FST. Collectively, these conclusions suggest that BMP antagonism is dictated by microenvironmental polarity in the desmoplastic invasion front.
FIGURE 5.5. Microenvironmental polarity of BMP inhibitors in CRC desmoplastic invasion front. (A-C) Mean expression scores of individual BMP1 determinants in tumor-budding and tumor-budding-free invasion fronts. Asterisks demonstrate statistically significant differences with Wilcoxon test (p<0.05). (D-E) Mean expression scores of individual BMP1 determinants across Q1-Q4 quartiles. Asterisks demonstrate statistically significant differences with either Wilcoxon test (matched contexts; p<0.05) or Mann-Whitney U-test (non-matched contexts; p<0.05). (F) IHC snapshots from one selected context. The preferential expression of HTRA3 in cancer-associated fibroblasts (CAFs) and the parallel expression of GREM1 and FST in cancer cells (i.e. microenvironmental polarity) are demonstrated. GREM1 signal is the lowest in this particular context, but is focally present in multiple neoplastic glands (i.e. low SI but high SP). Magnifications x200.
5.3.4. Expression Pattern of the BMP Inhibitor HTRA3 in Desmoplastic Invasion Fronts

Three diverse lines of evidence suggested that epithelial expression of HTRA3 followed a different pattern from the rest of the BMPIs that demonstrated a more conserved expression signature in CRC desmoplastic invasion fronts. First, eHTRA3 mean expression score shifted the pattern of the BMPI expression signature in the Q4 group of contexts, by incorporating itself in one of the observed ‘peaks’ (Appendix; Fig. D; red arrow). Second, this change was associated with significant increase in eHTRA3 levels of Q4 group of contexts, when compared to the rest of the quartiles (Fig. 5.5D.). Third, eHTRA3 failed to correlate with two other BMPI determinants (sHTRA3, Spearman’s rho=0.187, p>0.05; eFST, Spearman’s rho=0.196, p>0.05) (Fig. 5.4C.).

These discrepancies allowed us to hypothesize that the microenvironmental polarity of HTRA3 is prone to disruption in elevated levels of BMP antagonism and could be possibly related to biological phenomena in the desmoplastic invasion front, with most appealing speculation that of tumor-budding regulation. On this basis, we compared the mean expression scores of all six BMPI determinants between contexts designated as “invasion fronts” and those designated as “tumor-budding areas”, and found statistically significant increase of eHTRA3 expression levels (p=0.01; Mann-Whitney U-test) in the latter contexts (Figure 5.6A.).

As a proof-of-concept that HTRA3 changes microenvironmental polarity towards the epithelial compartment of tumors, a selected context (tGREM1=2, tHTRA3=10, tFST=3, tBMPI=15) demonstrating significantly increased HTRA3 immunoreactivity in tumor-budding cells (eHTRA3=6; sHTRA3=4) is shown (Fig. 5.6B.).
FIGURE 5.6. Epithelial expression of BMP inhibitor HTRA3 is differentially regulated between tumor-budding and tumor-budding-free desmoplastic invasion fronts. (A) Patterns of mean expression scores of BMP1 determinants between tumor-budding-bearing and tumor-budding-free contexts. Red arrow; monitoring of the eHTRA3 determinant. Statistically significant differences demonstrated with Mann-Whitney U-test (p<0.05). (B) IHC snapshots from one context demonstrating strong immunoreactivity of HTRA3 in the tumor budding subpopulation. Dashed lines: tumor buds. Magnifications x400.
5.3.5. Epithelial Expression of the BMP Inhibitor HTRA3 May Hold Pre-invasive Information for Tumor Budding Formation in Desmoplastic Invasion Fronts

Since tumor-budding is considered as an endpoint of cancer cell detachment from the proximal tumor invasion fronts, we sought to investigate whether eHTRA3 expression held any pre-invasive information in specific contexts for regulating it. To approach this question, we developed a classification model using logistic regression, to investigate whether there is an association between eHTRA3 expression and type of microenvironment in the desmoplastic invasion front (i.e. the binary outcome tumor budding-free or tumor budding-bearing microenvironment). An a priori goodness-of-fit test was performed, which demonstrated that the data were well-calibrated (p=0.339; Hosmer-Lemeshow test) for logistic regression analysis. Subsequently, logistic regression demonstrated statistically significant relationship between eHTRA3 expression score and tumor-budding-free microenvironment (p=0.025). ROC curves were conducted to evaluate the effectiveness of eHTRA3 in differentiating between tumor-budding-bearing and tumor-budding-free desmoplastic invasion fronts. The area under the curve (AUC) was 0.645 (Fig. 5.7A.) with a sensitivity of 89.1% and specificity of 20%, as depicted through cross-tabulation, showing that eHTRA3 is sensitive but not specific.

As a proof-of-concept, we demonstrated that high-levels of eHTRA3 were expressed in poorly-differentiated cancer cell cohorts positioned in the invasive margins (Fig. 5.7B.; yellow dotted line) in close proximity to the desmoplastic stroma. A striking difference of eHTRA3 expression between this cancer cell subpopulation (Fig. 5.7B.; yellow arrows) and the tumor core cells (Fig. 5.7B.; yellow asterisks) has been illustrated in many cases implying that these cells could probably demonstrate an “early tumor-budding phenotype”. Morphologically, these cells retained some level of cell-to-cell adhesion with the proximal invasion fronts (Fig. 5.7B.).
FIGURE 5.7. Epithelial expression of BMP inhibitor HTRA3 holds pre-invasive information for tumor budding formation in colorectal cancer desmoplastic invasion fronts. (A) ROC curve demonstrating the potential of the eHTRA3 expression levels in discriminating between tumor-budding-free and -bearing desmoplastic microenvironment. Area under the curve (AUC), p-values from logistic regression model and results from the Hosmer-Lemeshow (HL) test are described for each curve independently. (B) IHC of HTRA3 in two contexts, demonstrating high eHTRA3 expression in cancer cells lining the invasion front, predicting their potential for forming buds. Asterisks: negative tumor-core cells; dashed lines: tumor-host cell interface; arrows: HTRA3-positive cancer cell subpopulations of the invasion front undergoing detachment and possible early tumor budding formation. Magnifications x250.
5.3.6. Validation Experiments in an Independent Patient Cohort

Significant difference (p<0.05, Mann-Whitney U-test) of eHTRA3 expression between tumor budding-bearing and -free microenvironments was demonstrated in an independent cohort of 38 patients with CRC (Fig. 5.8A.). Expression scores from multiple contexts of the same case were combined in one mean HTRA3 expression score per case and correlation with tumor budding was, then, examined. The expression level of eHTRA3 was significantly increased (p=0.019, spearman’s rho=0.349), following an increase in the tumor-budding score at the level of the invasion front (Fig. 5.8B.). Logistic regression was conducted in the validation cohort, by additionally taking into consideration patient demographic data (i.e. age, gender, T-category, anatomical site, Crohn’s like reaction and marked-TILs) for potential confouders (Appendix; Table B). Interestingly, the expression of HTRA3 in cancer cells lining the invasion front could still significantly (p=0.042, AUC=0.731) predict the presence of tumor budding-free contexts with 64% sensitivity and 95% specificity (Fig. 5.8C.).
FIGURE 5.8. Validation experiments in an independent patient cohort. (A) Mean expression score of epithelial HTRA3 between tumor-budding-bearing and tumor-budding-free contexts in the validation cohort. Statistically significant difference demonstrated with Mann-Whitney U-test (p<0.05). (B) Correlation between tumor budding score and mean eHTRA3 expression level in each of the 38 cases of the validation cohort. (C) ROC curve in the validation cohort. Area under the curve (AUC), p-values from logistic regression model and results from the Hosmer-Lemeshow (HL) test are shown.
5.4. DISCUSSION

5.4.1. Introducing the Concept of “Context” in the Immunohistochemical Analysis

The immunohistochemical work presented in this chapter is based on the principle that desmoplastic interfaces expanding to different directions within the same tumor (i.e. in the same patient), might develop distinct molecular signatures. Such desmoplastic interfaces might be characterized by either presence or absence of EMT and tumor-budding formation, depending on the contextual framework, surrounding the cancer population. For this reason, the selection of desmoplastic CIFs for IHC analysis was based on the assumption that the surrounding context justifies their phenotypic differences. The assumption that cancer cells might receive signals from the ECM and the soluble factors trapped within (i.e. the context), to induce a particular phenotype has been very early conceived and adapted by Mina Bissell \textsuperscript{522}, and later on expanded to support a more generic model that context is the major driver of malignancy, since it regulates gene- and protein-expression in the adjacent tumor population, dictating its tumorigenic fate \textsuperscript{4, 523-526}. Therefore, our work examines the specific hypothesis that BMP suppression through extracellular BMP-antagonism might comprise one of the many prerequisites that characterize the EMT-promoting context of the desmoplastic interface. To which extent, these BMP-antagonists are sufficient and/or necessary to induce EMT, local invasion and tumor-budding formation in proximal cancer cells remains to be elucidated.

5.4.2. Characteristics of BMPI Expression Pattern in Desmoplastic Invasion Fronts

Two major characteristics jointly described the BMPI expression signature in CRC desmoplastic invasion fronts. First, the synchronous and coordinated stromal and epithelial expression of individual BMPIs in desmoplastic contexts, which demonstrated that all three of them contribute to increasing levels of BMP antagonism in such areas. Second, the presence of microenvironmental polarity in the BMPI pattern of expression, which was indicated through the preferential expression of HTRA3 in the stromal, and a parallel FST/GREM1 expression in the epithelial component of the investigated contexts.

Interestingly, we noted a synchronous and coordinated expression of GREM1, FST and HTRA3 in both stromal and epithelial compartments of desmoplastic cancers, which followed a consistent pattern as the levels of BMP antagonism elevated. As described earlier (chapter 1.3.6.), sporadic CRC often involves the disruption of the BMP pathway via a wide variety of mechanisms at the extracellular, receptor and/or downstream signal transduction level. Here, we found a coordinated increase of multiple extracellular
signals that may potentially induce BMP suppression. Among the most appealing speculations that could justify the coordinated expression of these signals in the CRC tumor microenvironment are: (I) the fact that such coordinated expression of BMPIs could serve as ‘safety valve’ to ensure the absolute BMP suppression \(^{345}\), and (II) the notion that many of these BMPIs act as complexes; interestingly certain BMP antagonist complexes might be even more effective than the affinity elicited by either of them alone \(^{234}\).

Moreover, by thoroughly investigating the BMPI signature pattern in desmoplastic contexts, we observed the presence of “microenvironmental polarity” in the expression pattern of these BMPIs. In particular, FST and GREM1 demonstrated preferential expression in the epithelial compartment of the tumors, while HTRA3 in the stromal compartment. Indeed, the BMP pathway has been documented to target both the epithelium and the mesenchyme, probably eliciting different responses in these lineages \(^{345}\). Interestingly, despite we found some expression of GREM1 in the tumor stroma, our most prominent observation was that of epithelial GREM1 expression, which comes in controversy with one study \(^{309}\), supporting an exclusive stromal GREM1 expression in squamous cell carcinoma, but at the same time comes in consistency with a different study \(^{527}\), supporting an epithelial expression of GREM1 in lung adenocarcinoma. Accordingly, HTRA3 expression in various contexts and types of cancer has a debatable microenvironmental polarity. For instance, HTRA3 is downregulated in endometrial cancers \(^{528, 529}\), but upregulated in thyroid cancers \(^{530}\). In our case, a significant shift towards the epithelial expression of HTRA3 was noticed with increasing levels of overall BMP antagonism, as well as with tumor budding. These observations may collectively support a speculative model, whereby BMP antagonists/inhibitors may possess a phenomenal plasticity of initiating or stopping their expression, to satisfy the needs of the epithelium or mesenchyme depending on the context. The exact mechanisms, regulating this microenvironmental shifting/polarity remain to be elucidated, although a very rapid and efficient shift could be easily attributed to epigenetic phenomena, as has been shown in the case of GREM1 \(^{531}\).

5.4.3. The Association of HTRA3 with Tumor-Budding

The expression levels of all three BMPIs demonstrated significant difference between microenvironments with and without tumor budding; however a correlation like this does not necessarily imply causation. More interestingly, the expression of HTRA3 in the epithelial compartment of the tumors demonstrated a significant predictive power to discriminate between these two types of microenvironments. Judging from the pattern of expression of eHTRA3 as observed in the examined contexts, it could be claimed that HTRA3 could serve as a rate-limiting step for the maintenance of a microenvironment devoid of tumor buds. In particular, the logistic regression model and the respective ROC curves suggested that as long as
the expression of the BMP antagonist HTRA3 is low in cancer cells lining the invasion front, there is an approximately 90% chance that the cancer will not form tumor-buds. However, to which extent the HTRA3-dependent BMP antagonism, as in the case of the other two BMP antagonists, is simply phenotypically coincidental or causatively linked with tumor budding formation remains to be elucidated.

Interestingly, tumor budding has been clearly attributed to malignant phenomena, such as EMT. Also, it has been suggested as an independent and reproducible factor of poor prognosis in CRC and other types of cancer, capable of stratifying those into risk groups, in a more meaningful fashion than the traditional TNM system. However, the assessment of the degree of tumor-budding is highly prone to laboratory- and observer-based biases. Therefore, well-standardized histological and IHC criteria for tumor budding evaluation remain to be determined. As such, the prognostic potential of HTRA3 as a predictive immunomarker, in the context of tumor-budding occurrence could be potentially regarded.

However, despite the fact that we observed significant predictive power of HTRA3 expression in identifying tumor budding-free microenvironments, this marker did not present with astounding sensitivity in defining the process. We believe that this was quite an expected and reasonable outcome, especially since tumor budding formation is a rather complicated process, possibly mediated by multiple EMT-regulating pathways, besides the BMP pathway. Interestingly though, our logistic regression model represents a very delicate means of showing that only a specific portion of CRC patients that eventually develop enhanced BMP antagonism (through aberrant HTRA3 secretion) will eventually form tumor buds in the invasion front. Therefore, we continue to foresee the presented logistic regression model as a statistical means for providing correlation insights, rather than a classification model to propose a predictive biomarker of the tumor budding process. Better description of the molecular features defining this CRC patient subgroup, along with a mechanistic explanation of how and why the polarity of HTRA3 expression in the tumor microenvironment is shifted should be further explored.

5.4.4. Validation Studies in the Independent Patient Cohort

A strong element in this study is the utilization of an independent patient cohort for validating the predictive potential of HTRA3 in discriminating between tumor-budding-free and -bearing microenvironments. By repeating logistic regression in this validation cohort, it should be noted that not only we were able to reproduce our finding, thus making our conclusions more concrete, but we have additionally taken into consideration a wide spectrum of confounding factors to correct our classification model, which are generally involved in disease progression.
For instance, important molecular pathways in CRC, such as the microsatellite stability status, should not be underestimated in the process of tumor budding formation, as these pathways have been documentedly related to the mechanism and nature of tumor invasion. It is known that CRC with MSI presents with lower aggressiveness, that may also be reflected in the macroscopic appearance of tumor configuration. This is also supported by the fact that RER+ tumors lose the EMT-promoting properties of TGF-beta signaling through mutational inactivation of TGFBR3, and such impairment of EMT may have an impact on the progression of the disease. Therefore, the currently described BMP antagonists could theoretically compensate for the missing EMT-promoting role of TGF-beta pathway in RER+ patients. This alternative strategy could help these cancers in promoting EMT and local invasion/tumor budding formation. This speculation remains to be explored. Unfortunately though, direct data on MSI status were not available in our patient cohorts, in the current study. However, based on previous knowledge, there is certain histopathology that is correlated with defects in mismatch repair mechanisms in CRC. In particular, Crohn’s-like lymphoid reaction and marked tumor-infiltrating lymphocytes (marked TILs), have been associated and very well-correlated with RER+ cancers. Information on these histologies was fortunately available in our verification cohort and these data were used for correcting respective biases in the logistic regression model.

Other factors corrected when the logistic regression model was repeated in the validation cohort were age, gender and anatomical site of the tumor. Indeed, the implication of these factors in the mechanism and nature of tumor development and progression should not be underestimated and neglected, as it is now known that they could invoke very diverse and important epigenetic profiles in the normal colonic mucosae, some of which could also be reflected in the neoplastic states. Interestingly, genome-wide methylation studies comparing the methylation status between different anatomical sites in the colon, revealed up to 8,300 differentially methylated loci between left and right colon, revealing the potential of an acquired functional specification among various anatomical sites in this organ. Such documented differences might also reflect to the broad diversity in secreted proteins, and provides a rationale for the observed “field cancerization” effect in CRC. To which extent, the expression of BMPI signature is affected by the field cancerization effect is a subject that should be addressed in future studies.

Despite the fact that the cohort size was adequate enough to compare expression of HTRA3 between the two types of contexts, it was unfortunately impossible to stratify these contexts, according to factors like age, gender, anatomical site and microsatellite status, because such process would generate strata with very small sample size and statistical hypothesis testing would be underpowered. At this particular point, we have used this patient information to simply perform correction of our logistic regression model from confounders. As such, this approach has definitely allowed us to support an association between HTRA3.
and tumor-budding formation, which seems to be independent of such factors. Future efforts should encompass validation in larger patient cohorts, where such stratification would be more feasible.

5.4.5. Speculations on Further Contributions of BMPIs to the Desmoplastic Invasion Front

The correlation of the expression of BMPIs with tumor budding and local invasion in the CRC desmoplastic interface does not necessarily imply causative association. As depicted, it is possible that BMPIs may be one part of the entire context that is essential for these phenomena, and other tumorigenic cytokines/growth factors may also be implicated in the process. Further experimentation to support a causative association of BMPIs and EMT needs to be done, to ascertain the model of BMP antagonism.

The BMP pathway, a signal transduction pathway that has been identified as an important barrier to intestinal tumor initiation \(^{345}\), has been shown to be disrupted via multiple mechanisms in sporadic CRC, including genomic mutations of SMAD4 and BMPR2, as quite early events during the development of the disease (i.e. at the adenoma-carcinoma sequence) \(^{360}\). Therefore, the possibility that many of the BMPI-expressing tumors included in our study are, in fact, already irreponsive to BMP signaling via another layer of BMP suppression (i.e. the genetic level), should not be neglected. In this likely scenario, there are two potential reasons for the observed contextual expression of the BMPI signature in the desmoplastic interface. First, the overrepresentation of these BMPIs might be an indirect consequence of the disturbed proteome of the many participants (cancer cells and CAFs) of the invasion front, without having an actual tumor-promoting, causative effect. Second, determinants of the BMPI signature might additionally regulate a diverse repertoire of biological responses in the desmoplastic invasion front.

Indeed, some of these BMPIs, besides acting as selective antagonists for BMP ligands, as in the case of GREM1 in particular, have been shown to bind to and activate phosphorylation of VEGFR2 in a BMP-independent fashion \(^{543}\), leading to the formation of VEGFR2/\(\alpha_\beta_3\)-integrin complexes and angiogenesis promotion \(^{544}\). Another one, FST, is less selective and although it heterodimerizes with BMPs with some affinity, the major pathway in the TGF-beta superfamily that it regulates, is the activin/inhibin pathway \(^{507}\), which is also known to be involved in CRC progression \(^{545-547}\). Finally, as explained earlier, HTRA3 participates in ECM remodeling in a manner that not only sequesters BMP ligands, but also many other members of the TGF-beta superfamily, reducing their bioavailability \(^{318}\). Consequently, the observed BMPI signature does not necessarily exert tumor-regulatory effects via BMP antagonism exclusively, since all the aforementioned pleiotropic effects have also been well-documented.
5.4.6. Conclusive Remarks

Emerging technologies, such as mass spectrometry-based proteomics and integrative bioinformatics \(^{166}\), have previously allowed us to propose the deregulation of multiple determinants of BMP antagonism in 
CRC \(^{453,514}\). In this chapter, we have successfully investigated the expression profile of this BMP signature in CRC patients through IHC, and illustrated the individual contribution of these BMPIs in the CRC desmoplastic interface.
CHAPTER 6
Contextual and Mechanistic Associations between Gremlin-1 Expression and Epithelial-to-Mesenchymal Transition

Sections of this chapter are published in Molecular Oncology:

Karagiannis GS, Berk A, Dimitromanolakis A, Diamandis EP. Enrichment Map Profiling of the Cancer Invasion Front Suggests Regulation of Colorectal Cancer Progression by Bone Morphogenetic Protein Antagonist, Gremlin-1. Molecular Oncology, 2013; 7: 826-839.


Copyright permission has been granted.
6.1. INTRODUCTION

Following the bioinformatic prioritization of bone morphogenetic protein inhibitors \(^{514}\), identified through proteomic characterization of an \textit{in vitro} coculture model, recapitulating the desmoplastic tumor-host cell interface in colorectal cancer \(^{453}\), we formulated a working hypothesis that this signature could provide context-dependent regulation of epithelial-to-mesenchymal transition, with subsequent local invasion and/or tumor budding formation.

To establish this link, we initially investigated the differential expression of GREM1, FST and HTRA3 in microenvironments with presence or absence of tumor budding, via immunohistochemistry. This analysis demonstrated: First, HTRA3 expression could adequately discriminate between these two types of microenvironment, and second, the BMP inhibitors/antagonists exerted microenvironmental polarity, i.e. HTRA3 was mainly expressed in the mesenchymal, while GREM1/FST were mainly expressed in the epithelial compartment of the desmoplastic tumor-host cell interface.

In this chapter, we extend our observations on the implication of BMPIs with various dedifferentiation phenomena at the tumor-host cell interface, such as the epithelial-to-mesenchymal transition biological reprogramming, irrespective of tumor-budding presence (which comprises a very unique mechanism of local invasion). To achieve that, we provide immunohistochemical evidence for contextual correlation of GREM1 expression with invasive cancer cell cohorts presenting with an EMT-related profile, and then, we expand our observations with \textit{in vitro} functional assays, whereby we establish a primitive causative link between BMP7/GREM1 and EMT/migration in colon cancer cell lines.
6.2. EXPERIMENTAL PROCEDURES

6.2.1. Cell Lines and Culture Conditions

HT29, SW480, SW620 and RKO colon cancer cell lines were obtained from the ATCC, Rockville, MD. These all belong to the sporadic CRC subtype and have the following characteristics: SW480 (CCL-228), Duke’s stage B, RER-; SW620 (CCL-227), Duke’s stage C, RER-; HT29 (HTB-38), Duke’s stage C, RER-; RKO (CRL-2577), Duke’s stage C, RER+. All cells were maintained in their favorable CM, according to the manufacturer’s instructions, in a modified atmosphere of 5% CO₂ in air in a humidified incubator at 37°C. All experiments were conducted within <5 passages from the initiation of all cultures.

6.2.2. Gene-Expression Analysis of EMT-related Genes, using Real-Time PCR

HT29 cells were seeded in triplicate in 12-well plates (DMEM, 10% FBS), to ~50% confluence. Under serum-free conditions, cells were stimulated for two days with rhTGF-beta (R&D Systems) or rhBMP7 (R&D Systems) or rhGREM1 (R&D Systems) or combination of any of the above, depending on the question asked. Following the termination of all in-vitro assays, cells were collected using non-enzymatic dissociation buffer (EDTA), and subjected to RNA extraction and cDNA synthesis, as described 184, 517, 518. RT-PCR was performed by using 1X SYBR green reagent (Applied Biosystems) and transcript levels of Snail-1, E-cadherin and N-cadherin were measured on a 7500 ABI System. The following 5'-3’ forward (F) and reverse (R) primer sequences were used:

**CADH1**- F: 5’-GCCTCCTGAAAGAGAGTGGAAG-3’, R: 5’-TGGCAGTGTCTCTCCAAATCCG-3’
**CADH2**- F: 5’-CCTAGAGTTTACTGCACATGAC-3’, R: 5’-GTAGGATCTCCGCCACTGATT-3’
**SNAI1**- F: 5’-TGCCTCCAAGATGCACATCCGA-3’, R: 5’-GGGACAGGAGAAGGGCTTT-3’
**GAPDH**- F: 5’-GTCTCCTCTGACTTTCAACAGCG-3’, R: 5’-ACCACCCTGTGTGCTGTAGCC-3’

6.2.3 Alkaline Phosphatase Assay

Alkaline phosphatase (ALP) activity was used to demonstrate active BMP signaling in various CRC cell lines, and was determined by para-nitrophenylphosphate (p-NPP) hydrolysis, as previously described 548, 549. In brief, following the termination of BMP7 stimulation experiments, cancer cells were washed twice with PBS and cell viability was assessed by exclusion of 0.5% Trypan Blue dye (data not shown). Approximately 500,000 cells in each condition were centrifuged for 5 min at 2,500 rpm, and their
supernatants were discarded. Then, the cell pellets were resuspended in 100 ml substrate buffer (10 mM diethanolamine, 0.5 mM MgCl₂, pH 10.5), and 1 mg/ml p-NPP was added. The mixture was incubated for 20 min at room temperature. The reaction was stopped with 50 ml 2 N NaOH. Optical density (OD) was measured at 405 nm. The optimization experiment showed that treatment with increasing doses of BMP7, increased the ALP activity of HT29, SW480 and SW620 in a dose-dependent fashion, whereas RKO cells served as negative control (Appendix; Fig. E). As such, we validated previous literature demonstrating that HT29, SW480 and SW620 cells could be used as an efficient in vitro model system for BMP signaling studies.

6.2.4. Morphological and Computer-Assisted Assessment of EMT

The similar experimental design with the one described above was reproduced for morphological assessment of EMT. HT29 cells were seeded in triplicate in 12-well plates (DMEM, 10% FBS), to ~50% confluence and under serum-free conditions, cells were stimulated with rhTGF-beta or rhBMP7 or rhGREM1 or combination of any of the above for two days. Evaluation of EMT incidence was performed through two independent methods, a more traditional operator-subjective and a computer-assisted one. For the former, three wells per experimental condition were fixed in 4% paraformaldehyde and stained with 0.5% crystal violet, as described. Five snapshots from each well (to a total of 15 pictures per experimental condition) were given to four blinded observers for independently assessing the percentage of cell population with epithelial phenotype. Score values of each experimental condition were combined for all snapshots from all different observers in one final mean score per experimental condition. For the latter, one-well per experimental condition was subjected to circularity index analysis, using ImageJ. Based on pixelled analysis of cell shape in a cell-by-cell basis, a perfect circle gives an index of 1.0, whereas the more polygonal, spindle-shaped or ellipsoid one shape is, the closer to 0.0 the index becomes. In brief, comparison of circularity indexes among the various experimental conditions was performed. For each condition, the circularity of approximately 120-150 individual cells was calculated, and a mean index was determined. Since HT29 cells tend to grow in colony configuration in vitro, at least three independent colonies/areas were selected in each experimental condition for evaluating the circularity of the cells. According to one study using the method, it is presumed that cells with small circularity index are more mesenchymal, thus representing populations that have possibly undergone EMT.
6.2.5. Under-the-Agarose Cell Migration Assay

Petri dishes were coated with fibronectin for 1 h at 37 °C. 1-cm-thick agarose gels, consisting of 2% low-melt agarose in DMEM, were cast into the dishes. After the gel was solidified, one large (~1 cm) well in the center, surrounded by four smaller (~0.5 cm) wells in the periphery of each Petri dish were punched, following the original protocol for the under-the-agarose cell migration assay. SW480/SW620 cells were stained with crystal violet, and treated for 24 h with serum-free DMEM. Following this process, 500 μL of approximately 5,000 cells/100 μl of serum-free DMEM were plated into the central cavity at the center of gel, and the surrounding holes were each filled with 200 μL of DMEM containing 10% FBS to create a chemotactic gradient. Then, specific factors under investigation, such as rhGREM1 and rhBMP7 were supplemented to the surrounding wells to create various experimental conditions (explained in the results section, in detail). Experimental conditions were applied for 48 h in a cell culture incubator, under standard conditions to allow cells to migrate, and following the termination of the experiments, the cells were fixed in 4% paraformaldehyde. Petri dishes were photographed and the obtained images were uploaded into ImageJ software for quantification of migratory behavior, as follows. First, a dashed circle was drawn in the periphery of the central cavity, to exclude cells that did not migrate from the central hole underneath the agarose. Second, the entire Petri dish surface was divided in four distinct quadrants, each one referring to the area of the chemotactic gradient created by the corresponding surrounding well. Based on pixelled analysis, the software allowed us to determine the total crystal violet intensity in the Petri dish, excluding the fluorescence shown in the central cavity. This score was determined as the “baseline intensity” for each image. To determine cell migration towards specific gradients, ImageJ was requested to provide the percentage of the baseline intensity that corresponded to the surface of each individual quadrant on one image at a time. In an optimization experiment (Appendix; Fig. F.), all four peripheral wells were filled with DMEM supplemented with 10% FBS to generate a chemotactic gradient and no additional motogens were utilized. Expectedly, approximately 25% of the total cell population was found to be migrating towards each independent quadrant (Appendix; Fig. F.), suggesting that differences in real experimental setups could be potentially attributed to the motogens added in certain quadrants.

6.2.6. Retrieval of Archived Immunohistochemical Cases and GREM1-EMT Correlation Studies in Selected Immunohistochemical Contexts

We utilized archived IHC slides from the experiments conducted in the 30-CRC-patient cohort, presented in chapter 5 to demonstrate additional evidence for contextual correlation of GREM1 expression with desmoplastic interfaces undergoing EMT and localized invasion, irrelevant of tumor-budding formation.
To achieve that, we followed a similar principle as in the previous chapter, by selecting areas in a context-dependent fashion. In all these investigations, GREM1 was scored as described in Chapter 5, and therefore, details on the scoring system are not repeated here.

Two studies were performed with regards to linking the contextual expression of GREM1 with dedifferentiation (i.e. EMT-related phenomena) in desmoplastic interfaces. The contexts selected in each case were subdivided in “EMT-pertaining” and “EMT-quiescent” to demonstrate an active or non-active dedifferentiation process. “EMT-pertaining” contexts were defined according to the following criteria: (I) cancer cells positioned at the periphery of the cohorts presented with nuclear accumulation of beta-CAT, while those in the core of the cohorts with membranous, AND (II) the expression levels of the epithelial tight-junction marker OCLN may or may not present with similar gradation; however, a progressive loss in the expression of OCLN in these cancer cells, when compared to control areas (i.e. normal colonic mucosa) and earlier invasion areas should be present. A failure to fulfill either of these two criteria categorized the contexts as “EMT-quiescent”. Representative figures, which clearly portray the differences between cells expressing membranous or nuclear beta-catenin expression, as well as the occludin expression gradient, are shown (Fig. 6.1.).

In the first study, we selected contexts of deep cancer invasion, also known as “intramuscular invasion fronts” (IIFs), which present with the typical appearance of finger-like projections, penetrating the smooth muscle, below the submucosal layers (i.e. muscularis externae). A total of 28 IIF contexts were found during our screening in these archived cases, for which the co-assessment of beta-catenin and OCLN successfully categorized them into 21 “EMT-pertaining” and 7 “EMT-quiescent” contexts.

In the second study, we selected submucosal contexts of isolated, cancer cell aggregates with absence of glandular or even pseudoglandular architecture, also known as tumor-nests (TNs). It is believed that such multicellular aggregates might have possibly detached from a larger cancer cell cohort/collective invading the stroma, through a poorly understood mechanism of pathogenesis. The most attractive speculation, though, is that of a stiffness-mediated disconnection through intense mechanical and reciprocal interactions of cancer collectives with the tumor-associated stroma. The fate of these smaller tumor nests is currently unknown, despite preliminary evidence has been linking them with a passive, invasion-independent metastatic pathway, initiated by intravasation of tumor nests enveloped by endothelial cells of sinusoidal vasculature. A total of 42 TNs were identified during our screening in these archived cases, and they were further subdivided into 31 “EMT-pertaining” and 11 “EMT-quiescent” contexts, using the co-assessment of beta-catenin and OCLN, as described in the case of the IIF contexts.
6.2.7. Statistical Analysis

Statistical differences were assessed with non-parametric approaches, in particular the Jonckheere-Terpstra test for multiple (>2) independent groups and the Mann-Whitney U-test for 2 independent groups. Statistical significance was generally shown at the 0.05 level, but where appropriate, the p-values were adjusted for multiple-hypothesis testing, using Holms correction. Data are shown as means with their standard errors. The SPSS (version 20) software was used for these analyses.
FIGURE 6.1 Occludin (OCLN) and beta-catenin immunohistochemistry. (A) Assessment of OCLN dedifferentiation gradient across the various compartments of primary CRC. All magnifications X400, except left figure X210. (B) Immunohistochemical differences between beta-catenin nuclear- (N-) and membrane- (M-) expressing cancer cells. The two figures on the right demonstrate Ki-67 and laminin-alpha-3 expression in the same area as the figure on the left. All magnifications X200, except left figure X240.
6.3. RESULTS

6.3.1. Contextual Correlation of Peritumoral GREM1 Expression with Selected Contexts Exhibiting an EMT-Pertaining Phenotype

As explained earlier, IIFs are characterized by the massive and collective invasion/migration of cancer tissue in muscularis externae in an interdigital pattern (Fig. 6.2A.). A careful observation of the 28 selected IIF contexts revealed that the cancer cell cohorts were circumscribed by ECM-enriched sheaths, possibly of desmoplastic nature, of low-to-medium thickness. We concluded that these ECM-enriched sheaths were of myofibroblastic/CAF structure, as they retained the typical signature of such microenvironments, i.e. they were positive for the epithelial-myofibroblastic basement membrane protein LAMB1, positive for the smooth muscle/myofibroblastic marker alpha-SMA and negative for the smooth muscle marker h-CALD (Fig. 6.2B.). Therefore, these invasive cancer cell cohorts do not come in direct contact with smooth muscle tissue, but they seem to migrate alongside CAFs. Two additional hallmarks of desmoplastic reaction were present in these myofibroblastic sheaths. First, the myofibroblastic cells focally expressed the marker of proliferation Ki-67 (Fig. 6.2B.). Indeed, it is known that desmoplasia is not only characterized by abundant deposition of ECM components but also by the molecular recruitment and proliferation of myofibroblasts around cancer cells. Second, there was a prominent population of cluster of differentiation 31 (CD31) expressing cells (Fig. 6.2B.). This finding may suggest that increased angiogenesis is invoked within these migratory regions, since we have already extensively demonstrated in chapter 1.1.2.3 that CAFs may secrete a plethora of proangiogenic factors. In addition, the scattering pattern of these CD31+ cells implies the deployment of a de novo angiogenic program in this area, rather than invasion of such enormous cancer cell cohorts in pre-existing blood vessels. GREM1 expression in these desmoplastic sheaths was variable but detectable in most cases (Fig. 6.2B.). Following the characterization of these peritumoral desmoplastic sheaths in the IIF contexts, we then proceeded to compare the expression levels of GREM1 between EMT-pertaining and EMT-quiescent contexts. Interestingly, GREM1 expression was significantly higher (p<0.05; Mann-Whitney U-test) around the EMT-pertaining contexts compared to the EMT-quiescent ones (Fig. 6.2C.).
FIGURE 6.2. Association of GREM1 expression with EMT in intramuscular invasion fronts. (A) Illustration explaining the selection of IIF contexts. The cartoon on the left demonstrates the design of selection, while the figure on the middle demonstrates an example of IIF (beta-catenin staining). The demarcated yellow-squared area is further shown magnified with both beta-catenin and OCLN staining on the far right. (B) Set of figures of the selected context shown in (A) indicating the staining of each of the specified markers (panel above) and a cartoon illustration (panel below), which demonstrates an overlap of the expression of the corresponding protein with that of GREM1. GREM1 is expressed in the thin desmoplastic, peritumoral sheath, as the cancer cell cohorts invade into the muscularis externae. This sheath is LAMB1-positive, alpha-SMA-positive and h-CALD-negative (upper panels). The desmoplastic sheath is especially distinguished easily in the h-CALD panel, as a thin layer between cancer and muscle cells. Also, the cancer cells and their peritumoral stroma show some variable degree of proliferation (Ki-67 staining), while the peritumoral sheath is focally immunoreactive with CD31, indicating increased angiogenesis. (C) Contextual association of GREM1 expression with EMT. The EMT-pertaining cancer cell cohorts of the IIFs demonstrate prominent EMT gradation with the following characteristics: nuclear staining of beta-catenin in the periphery with parallel membranous expression in the core of the cells (upper panel). The cartoon schematic shows this gradation of the EMT process, as implied from the differential localization of beta-catenin staining in such cancer cell cohorts (middle panel). The graph demonstrates statistically significant difference (asterisk demonstrates p-value of <0.05; Mann-Whitney U-test) of GREM1 expression in the desmoplastic sheaths between EMT-pertaining and EMT-quiescent IIFs.
As explained earlier, TNs are characterized isolated multicellular aggregates, which have possibly detached from larger cancer cell cohorts invading the stroma (Fig. 6.3A.). These TNs might be associated with a passive, invasion-independent metastatic pathway, initiated by their intravasation after getting enveloped by endothelial cells of sinusoidal vasculature. A careful observation of the 42 selected TN contexts revealed that the cancer cell cohorts were circumscribed by ECM-enriched envelopes, whose structure was possibly of a combination of desmoplastic and sinusoidal nature, since they were positive for LAMB1, and remarkably circumscribed by very thin alpha-SMA-positive and h-CALD-positive sheath (Fig. 6.3B.; yellow arrows). Most of these TNs (N=31) presented with very prominent OCLN and beta-catenin gradients (Fig. 6.3C.; upper panels), however a few (N=11) were also categorized in the EMT-quiescent phenotype. Interestingly, stromal GREM1 expression around these TNs was mild and variable in most cases, but focal expression in a few cancer cells in the periphery of these nests was also prominent in a few cases (Fig. 6.3B.). Following the characterization of these peritumoral envelopes in the TN contexts, we then proceeded to compare the expression levels of GREM1 between EMT-pertaining and EMT-quiescent contexts. Interestingly, GREM1 expression was significantly higher (p<0.05; Mann-Whitney U-test) around the EMT-pertaining contexts compared to the EMT-quiescent ones (Fig. 6.3C.).
FIGURE 6.3. Association of GREM1 expression with EMT in isolated tumor nests. (A) Illustration explaining the selection of TN contexts. The cartoon on the left demonstrates the design of selection, while the figure on the middle demonstrates an example of TN (beta-catenin staining). The demarcated yellow-squared area is further shown magnified with both LAMA3 and Ki-67 staining on the far right, to clearly demarcate and characterized the epithelial compartment. (B) Set of figures of the selected context shown in (A) indicating the staining of each of the specified markers (panel above) and a cartoon illustration (panel below), which demonstrates overlap of the expression of the corresponding protein with that of GREM1. GREM1 is mildly expressed in stromal cells comprising the envelope and focally by cancer cells at the periphery of these TNs. The TN envelopes are LAMB1-positive, and remarkably have very thin alpha-SMA-positive and h-CALD-positive layer around (yellow arrows). (C) Contextual association of GREM1 expression with EMT. The EMT-pertaining cancer cell cohorts of the TNs demonstrate prominent EMT gradation with the following characteristics: focal nuclear staining of beta-catenin in the periphery with parallel membranous expression in the core of the cells (right panel) and progressive loss of OCLN from the core to the periphery of TNs (left panel). The cartoon schematic in the low-left corner shows the gradation of the EMT process, as implied from the panels above. The graph in the low-right corner demonstrates statistically significant difference (asterisk demonstrates p-value of <0.05; Mann-Whitney U-test) of GREM1 expression in the TN envelopes between EMT-pertaining and EMT-quiescent TNs.
6.3.2. The BMP7-GREM1 Axis Promotes the EMT Phenotype in Colon Cancer Cells *in vitro*

Phenotypic EMT assays were performed in HT29 colon cancer cells, since their ability to grow in colonies and their relatively well-differentiated phenotype could provide safe assessment of an active EMT process. In addition, this cell line has been used for EMT studies in the past, which provided guidance for the current investigations. Morphological assessment of the epithelial/mesenchymal phenotype ratio in BMP7/GREM1 stimulation assays revealed that the addition of GREM1 alone, or in combination with BMP7, was able to cause some degree of disruption of cell-to-cell adhesion in HT29 colonies, and promoted the mesenchymal phenotype, especially in cancer cells that lined the periphery of the HT29 colonies (Fig. 6.4A.). Quantification of this experiment with blinded observers demonstrated a significant decrease (p<0.05, Jonckheere-Terpstra test) in the percentage of cells with epithelial phenotype, after treatment with GREM1, in a dose dependent fashion (Fig. 6.4B.). The effect kept being significant even with the parallel administration of BMP7 at concentrations of 50 ng/mL (Fig. 6.4B.). However, the addition of BMP7 alone did not significantly affect the epithelial phenotype or ability of HT29 cells to form their typical cohorts/colonies (Figs. 6.4A. & 6.4B.), an observation that further strengthens the notion that BMP7 may promote the integrity and maintenance of the epithelial phenotype in CRC. Despite subjective to a certain extent, the results presented in Fig. 6.4B were further confirmed through computer-assisted determination of cell circularity (Fig. 6.4C.), as previously described. In particular, the mean circularity index of HT29 cells was significantly reduced (p<0.05; Jonckheere-Terpstra test), following either GREM1 treatment or combined GREM1/BMP7 treatment, and remained intact (p>0.05, Jonckheere-Terpstra test) following treatment by BMP7 alone (Fig. 6.4C.).
FIGURE 6.4. Morphological assessment of EMT phenotype regulation by the BMP7/GREM1 axis in vitro. (A) A representative snapshot from each experimental condition. All magnifications X400. (B) Mean percentage of HT29 cells retaining an epithelial-like phenotype in the EMT assay. Each bar refers to the mean percentage determined for each experimental condition demonstrated in (A). Asterisks denote statistical significance, p<0.05, Jonckheere-Terpstra test. (C) Mean circularity index of HT29 cells in the EMT assay. Each bar refers to the mean circularity index determined for each experimental condition demonstrated in (A) independently. Asterisks denote statistical significance, p<0.05, Jonckheere-Terpstra test. The panel at the end shows the treatments in each experimental condition and refers to both (B) and (C) graphs.
6.3.3. The BMP7-GREM1 Axis Shifts the E/N-Cadherin Switch and Induces the Overexpression of the EMT-Promoting Transcription Factor and EMT marker Snail-1

The phenotypic EMT-related changes following GREM1 treatment as well as GREM1-dependent suppression of BMP7 in HT29 colon cancer cells (i.e. chapter 6.3.2. and figures therein) was further shown through the induction of specific EMT markers at the gene expression level. In particular, we determined the transcription status of two independent machineries, first the induction of the cadherin switch, and second the induction of EMT-promoting transcription factors, such as Snail-1\textsuperscript{99, 469, 470}. For reproducibility purposes, the same experimental conditions as the ones described in the phenotypic assays were kept, during the evaluation of these EMT markers in HT29 cells. In addition, during assay optimization, we noted that stimulation with TGF-beta (5 ng/mL), a cytokine known to induce EMT in CRC and other cancers\textsuperscript{560}, was able to induce the cadherin switch (data not shown), thus confirming previous studies\textsuperscript{559, 561, 562} that have established the HT29 cell line as a model system for studying EMT induction. We then noticed that the stimulation of HT29 cells with BMP7 alone caused a significant (p<0.025, Holms-corrected Jonckheere-Terpstra test) shift of the cadherin switch towards the non-EMT phenotype, in a dose-dependent fashion (Fig. 6.5A.). Moreover, the stimulation with GREM1 alone caused a small, but significant (p<0.025, Holms-corrected Jonckheere-Terpstra test) shift towards the EMT phenotype, in a dose-dependent fashion (Fig. 6.5A.), confirming our notion that the BMP7/GREM1 axis could regulate EMT in the CRC invasion front. However, the fact that the cadherin switch was not altered severely by GREM1 stimulation alone (E-cadherin was decreased by ~1.3-fold, while N-cadherin was increased by ~1.8 fold with a single dose of 1000 ng/mL GREM1) prompted us to speculate there may not be high enough levels of endogenously-secreted BMP7 to be efficiently antagonized in this in-vitro setting. Let alone, most studies that have induced EMT in HT29 cells have observed significant results in ~8 days post stimulations, and our experiments were terminated in 2-days post-treatment. To circumvent this issue, we repeated GREM1 stimulation in this setup, but included co-stimulation with 50 ng/mL of BMP7 in all conditions. This experiment allowed us to better observe the GREM1/BMP7 effect on the cadherin switch, and again, a statistically significant (p<0.025, Holms-corrected Jonckheere-Terpstra test) shift of the cadherin switch towards the EMT phenotype was observed (Fig. 6.5A.). Furthermore, our observations on the gene expression changes of Snail-1, an EMT-related transcription factor, were in absolute concordance with the expression pattern of the mesenchymal N-cadherin (Fig. 6.5B.), further confirming our notion that GREM1 may induce EMT in HT29 cells, by either antagonizing endogenous or exogenous BMP7, among other possible mechanisms.
FIGURE 6.5. Regulation of EMT markers by the BMP7/GREM1 axis in HT29 colon cancer cells in vitro. (A) Mean gene expression levels of E-cadherin (E-CAD) and N-cadherin (N-CAD). The expression levels follow the stimulation of HT29 cells with varying doses of BMP7 (0.5, 5 and 50 ng/mL), GREM1 (10, 100 and 1000 ng/mL) or both (BMP7 at 50 ng/mL; GREM1 at 10, 100 and 1000 ng/mL). All bars demonstrate relative values to the non-treated control. Asterisks denote statistical significance, p<0.007, Jonckheere-Terpstra test. (B) Mean gene expression levels of Snail (SNAI1). The expression levels follow the stimulation of HT29 cells with the exact same experimental conditions shown in (A). All bars demonstrate relative values to the non-treated control. Asterisks denote statistical significance, p<0.007, Jonckheere-Terpstra test.
6.3.4. The BMP7-GREM1 Axis Promotes Colon Cancer Cell Migration in vitro

A feature involved in the potential of GREM1 to induce localized invasion in desmoplastic CIFs, in an in vivo context, could be the ability to enhance a migratory program, besides mediating EMT (as we demonstrated in chapters 6.3.2 and 6.3.3). Here, we provided some preliminary evidence to test the accuracy of this speculation in a simplistic in vitro setting, by performing the traditional under-the-agarose cell migration assays. Such assays depend on the generation of chemotactic gradients of motogens that are diffused in agarose gels. Unfortunately, we have been unable to perform these assays with HT29 cells and maintain constancy in our results, since HT29 cells could not migrate away from the initial seeding position, but they rather formed colonies, and migrated slowly through invasive growth. In contrast, SW480/SW620 cells that have been used widely in various cell migration assays in the past, represented attractive candidates for this assay. During the control experiment, where no motogenic gradients were applied, an equal distribution of migratory cancer cell populations in each quadrant was observed. Computer-assisted quantification indicated that ~25% of the total crystal violet staining intensity of the entire Petri dish was distributed in each quadrant, which demonstrated that SW480 and SW620 cells could generally migrate evenly (Appendix; Fig. F).

Following this optimization, we proceeded to investigating the influence of chemotactic gradients of GREM1 and/or BMP7 on the migration of both SW480 and SW620 cells. The tested motogen, in this case rhGREM1, was initially diluted in wells positioned in the periphery of Petri dishes and over time the developing gradients affected the migratory behavior of cancer cell populations residing in the central cavity. To monitor migration of the cancer cell cohort under the influence of the GREM1 motogen, we fixed and stained the cancer cells with crystal violet after the termination of all experiments, and subsequently used imaging software to calculate the percentage of cancer cells positioned on each quadrant corresponding to one of the four individual gradients. Our initial observations on this setup revealed that more than 70% of SW480/SW620 cells migrated towards an rhGREM1 gradient (2 mg/mL) (quadrant B), while lower populations (<10%) migrated towards the absence of such gradient (quadrant A), or towards a gradient of rhBMP7 (500 ng/mL) (quadrant C) (Fig. 6.6A.). Especially the latter is an interesting observation, because it demonstrates that BMPs not only decrease cancer cell migration as also suggested by the literature, but may also act as repellents of cancer cell motility. Finally, only a small SW480/SW620 population (10%) migrated towards a gradient consisting of admixed rhBMP7 and rhGREM1 (quadrant D) (Fig. 6.6A.). This may possibly suggest that when rhGREM1 gradient is disrupted by addition of rhBMP7, then under-the-agarose cancer cell migration is abrogated in this specific quadrant, possibly through reduction of rhGREM1 bioavailability from covalent binding and complex formation with rhBMP7. However, this interaction, if indeed occurs, remains to be elucidated.
Subsequent experiments revealed that SW480/SW620 cancer cells migrated towards GREM1 gradients in a dose-dependent fashion. Specifically, approximately 50% of the cells migrated towards quadrant D, the strongest gradient (4 mg/mL of rhGREM1) of all, approximately 30% towards quadrant C (2 mg/mL of rhGREM1), while the remaining 20% of cells were collectively found on quadrants A and B, which held no or very weak (1 mg/mL of rhGREM1) gradients, respectively (Fig. 6.6B). Along with data presented in Fig 6.6A, these data collectively indicate that the observed migratory behavior of SW480/SW620 cells in quadrant B could indeed be GREM1-dependent.

Overall, we believe that data presented in this section are in concordance with those presented in the previous chapters 6.3.2/6.3.3, in a fashion suggesting that GREM1 may promote EMT and motility of cancer cells through BMP7 suppression. These in vitro experiments may provide, if interpreted with caution, an in vivo analogy that paracrine secretion of GREM1 could antagonize BMP7 in the desmoplastic CIF and possibly exert tumor-promoting effects, such as EMT and enhanced motility, thus supporting localized invasion of cancer cells lining the invasive margins.
FIGURE 6.6. GREM1 stimulates migration of SW480/SW620 cells in vitro in a BMP7-dependent manner, as shown through under-the-agarose cell migration assays. Panels on the right describe in detail the gradients induced in each of the quadrants A-D. The corresponding graphs below describe the crystal violet intensity found in each individual quadrant, as percentage of the total fluorescence that was detected in the entire plate. The values are mean percentages of intensities from three independent replicates. The figures show one representative replicate from each cell line.
6.4. DISCUSSION

6.4.1. Contextual Association between GREM1 and EMT: Immunohistochemical Evidence

In this chapter, we provided IHC evidence of contextual correlation between GREM1-expressing stromal cells and EMT-undergoing cancer cells in CRC desmoplastic interfaces. To demonstrate this, we again utilized a context-by-context logic in our analysis, as described in chapter 5, based on the principle and definition of “tumor context”, established previously by Mina Bissell 4. In this study, EMT was assessed by co-evaluating strong nuclear beta-catenin and progressive loss of the tight junction marker occludin, two quite prominent hallmarks of the process 99, in sequential sections of the exactly same tumor areas. However, it should be noted that IHC is a snapshot of the constantly evolving and dynamic process of carcinogenesis in the diseased tissue. As such, the here-reported EMT-pertaining phenotype does not necessarily coincide with a fully-acquired mesenchymal phenotype in these cancer cells, but it rather implies that their gene- and protein-expression machinery is simply skewed in reprogramming their phenotypic properties towards more mesenchymal ones.

Colorectal cancers arising from the mucosal epithelium may penetrate the underlying tissues and in form of collective migration breach very deeply into the various tissue layers, while they may sometimes even reach the serosa and invade and metastasize in the peritoneal lamina 520, 564. The presence of such areas of deep invasion is also considered as an adverse prognostic factor for CRC 520. During histopathology, it is very frequent to note that the downward cancer cell collectives have already breached the muscularis externa 564. In general, the smooth muscle layers in the healthy colonic tissue depict important biological barriers for CRC progression, and it is the breaching of the muscularis mucosae the key event that discriminates colorectal cancers into stage I and stage II 520. Because smooth muscle cells are prominent biological barriers, it is not very frequent to observe tumor buds in those areas either 153, but cancer cells tend to move in an interdigital pattern, disrupting tissue architecture along the way 564. Here, we investigated these “intramuscular invasion fronts” and reported that most of them had an EMT-pertaining phenotype in their invasive edges with parallel GREM1 expression in the surrounding desmoplastic stroma. Concordantly, these cells could have adapted the orientation of the “leading” or “tip” cells, according to the model of collective migration described by Rorth 143. In this model, an active EMT phenotype is deployed by the edge cells to assist in the migration process, while cells at the tumor core are keeping a purely epithelial phenotype, with strong anchoring (i.e. tight junctions) to the edge cells while they are getting passively pushed. To which extent, a GREM1-dependent EMT induction is indeed maintained in these intramuscular invasion fronts is a subject that remains to be elucidated.
Our immunohistochemical observations additionally demonstrated focal GREM1 staining around areas of vascular invasion, designated as TN contexts (Fig. 6.3.). The invasive behavior of cancer cells is generally essential during the metastatic cascade, not only for allowing cancer cells to breach into the various submucosal components, but also for enabling them to penetrate through the vascular wall barrier (a.k.a. intravasation) at either the primary or secondary site of growth. An alternative model, in which tumors could gain access to the host vasculature through an invasion-independent mechanism, has been previously proposed. Specifically, a murine model of mammary tumor metastasis was developed, whereby it was shown that tumor nests (TNs) became progressively surrounded by sinusoidal blood vessels and entered the circulation as endothelium-coated tumor cell emboli. Similar encoating of endothelium around tumor emboli had been previously described, but the mechanism and clinical significance of this phenomenon has only been recently investigated. A common feature of the vasculature in the primary tumors carrying endothelium-coated tumor emboli is the presence of sinusoid structures surrounding TNs, which was also observed in our study by selecting these contexts. These re-organized tissue structures consisting of TNs and the sinusoidal envelop presumably become the embolus ‘unit’, which may either be disseminated or force the collapsing of the vessel. Therefore, this pathway may depend not on tissue destruction by cancer cell invasiveness but on remodeling of tissue architecture through tumor-stroma interactions. Our IHC studies on the interface of these interactions indicated that GREM1 is present in and around this remodeled endothelium, further suggesting that it could potentially induce EMT and/or participate in the intravasation of the TNs within the vessels. Despite the fact that we have not as yet provided any causative association of EMT or vascular remodeling, GREM1 represents a very promising mediator of this phenomenon, for yet another reason. It has been shown that GREM1 is associated with neoangiogenesis/neovascularization in a BMP-independent fashion, as GREM1 is capable of binding to, and activating, phosphorylation of VEGFR2 on the surface of endothelial cells.

6.4.2. Functional Association between GREM1 and EMT/Migration: Preliminary Evidence

Here, we also provided preliminary functional evidence that the BMP7/GREM1 axis could potentially regulate EMT and migration of colon cancer cells in vitro. Both the EMT and the migratory machinery comprise two important cell biological programs for invasion and metastatic dissemination of cancer cells, and they are both regulated by extensive, partly-overlapping gene- and protein-expression cellular circuitries, involving some prominent pathways, i.e. TGF-beta, Wnt, Notch and Hedgehog. These two biological programs could be deployed in strict coordination with each other to regulate local invasion and tumor budding formation. First, EMT could be responsible for promoting the mesenchymal...
and spindle shape in the cancer cells of the invasion front, and could also release them from the firm anchoring with the tumor core cells, by disrupting tight-, gap- and adherens-junctions. Subsequently, these cells could deploy a motility program to further penetrate and migrate into the desmoplastic stroma.

Despite EMT is sufficient, it does not seem to be necessary (as enhanced motility is) for tumor progression, since there is now robust evidence demonstrating that many tumors metastasize through collective migration only, and local invasion does not occur in the form of EMT and/or tumor budding at the level of the invasive front. For instance, Gavert N. *et al.* (2005) described specific subpopulations of colon cancer cells that could invade, migrate and metastasize without necessarily undergoing EMT and/or tumor-budding formation. In another study, the expression of podoplanin, a small mucin-like protein, could promote local invasion and collective migration in breast cancer cells *in vitro* and *in vivo*, in a filopodia-dependent mechanism, again in the absence of EMT. Therefore, over a series of reported studies, the absence of EMT did not attenuate the metastatic cascade and did not pose a barrier for invasion and metastasis; the latter was not achieved through tumor-budding formation, but through a rather aggressive and migratory CIF.

In these preliminary experiments, EMT was functionally assessed with both phenotypic and specific marker-monitoring assays. It is generally very difficult to perform morphological assessment of EMT, since this phenomenon, in fact, describes the dynamic and progressive transition of the differentiation status from an epithelial-like into a mesenchymal-like state, and it is not a *bimodal* phenomenon, as was previously believed. Significant issues are raised when cancer cells are already quite mesenchymal in shape (i.e. SW480/SW620), and as such phenotypic assessment is not quite possible using subjective methods (i.e. blinded observers). Additionally, morphological hallmarks like detachment of individual cells in cell lines where the cells tend to make colonies/multicellular aggregates (i.e. HT29) may facilitate EMT interpretation, as in our case. However, it should always be kept in mind that *in vitro* changes might have altered the morphology of cancer cells over a number of passages, and EMT evaluation might not reflect an actual *in vivo* setting. For this reason, phenotypic assays should always be accompanied by gene- and or protein-expression changes in specific markers that have been well-documented to signal an EMT event. Our *in-vitro* EMT assays were based on measuring the gene-expression levels of the EMT-initiating transcription factor *snail* and the well-established *cadherin switch*. Either directly (first response targets) or indirectly (second response-targets), all three markers have been explicitly and undeniably associated with EMT, and have been extensively used in cancer research. Using a combination of these assays, we provided preliminary evidence that BMP7 could promote and maintain HT29 differentiation (by inhibiting EMT or promoting the reverse program i.e. MET), and that GREM1 could overcome and reverse this effect.
Despite clear evidence from developmental studies, the modulation of differentiation (re)programming (and EMT) in the cancer setting has been a subject of major speculation for the BMP pathway and the respective antagonists \(^{233, 276, 581}\). Such debate was raised mainly due to the documented difficulty of studying the BMP pathway in cancer progression, as there have been the following: First, the broad spectrum of BMP ligands, receptors and antagonists in the tumor microenvironment which may compensate for one another and which may or may not be interconnected under the same downstream pathway; second, the wide variety of canonical/non-canonical downstream effects of the BMP pathway; third, the regulation of their expression by multiple transcription factors and co-regulators; fourth, the complicated intracellular interactome of almost all signaling mediators from the receptor level down to the transcription factor/co-regulator level \(^{233, 276, 581}\). This complex signaling relay has been thoroughly described in chapter 1.3.

### 6.4.3. Conclusive Remarks

In this chapter, we decided to focus our proof-of-concept studies on BMP antagonist GREM1 exclusively, leaving both HTRA3 and FST besides, for the following reasons: (I) GREM1 is a highly-selective antagonist of BMP2/4/7, and it does not bind to and inhibit other members of the TGF-beta superfamily \(^{276}\). On the contrary, FST is primarily an antagonist of the activin/inhibin pathway and can only bind with very low efficacy to some BMP ligands \(^{507}\); (II) For IHC purposes, GREM1 had the best quality controls among all BMPis tested, since this protein is known to be expressed in pericryptal myofibroblasts of healthy colonic mucosae/crypts, whereby it maintains a stem cell niche \(^{521}\); (III) There is commercial availability of recombinant GREM1 protein, which could inhibit the BMP2 and BMP7 signaling cascade in mature osteoblasts, according to the manufacturer, making rhGREM1 suitable for testing our hypothesis in an *in vitro* setting; (IV) Unlike GREM1/BMP where the mediated antagonism is selective and achieved through covalent binding and heterodimerization \(^{582}\), the HTRA3/BMP relationship is different as it is based on ECM proteolysis and extracellular BMP sequestration \(^{318}\). Because such interaction requires more complicated *in vitro* design, here, we decided to simply study the BMP/BMPi interactions, like the ones between GREM1 and BMP7.

Overall, this chapter examined the cotextual and mechanistic associations between the expression-function of the BMP antagonist GREM1 and EMT/migration. Since the deployment of these two interrelated biological programs could be causatively associated with local invasion and tumor budding formation, we may conclude that GREM1-dependent BMP antagonism at the level of the invasion front may represent a potential strategy for the completion of the metastatic cascade.
CHAPTER 7

Discussion and Future Directions
7.1. SUMMARY AND CONCLUSIONS

7.1.1. Brief Overview

This thesis has successfully described a comprehensive molecular portrait of the tumor-host cell interface of desmoplastic cancers, out of which has also proposed a group of extracellular bone morphogenetic protein inhibitors/antagonists, as potential mediators of local invasion and tumor budding formation. In specific, the expression of GREM1, FST and HTRA3 was further investigated in desmoplastic lesions of colorectal cancer patients, and a paracrine axis between GREM1/HTRA3 and dedifferentiation, such as epithelial-to-mesenchymal transition, was subsequently suggested, as an impending strategy for inducing local invasion and tumor budding formation at the invasion front.

We first proposed an innovative mass-spectrometric approach based on the emerging field of “cancer secretomics”, to characterize the molecular portrait of the cancer invasion front, based on the fact that it is mainly desmoplastic in many types of cancer, such as the gastrointestinal ones (objective #1). The proteomic signature generated provided a framework for subsequent meta-mining, using innovative and integrative bioinformatic tools, based on the principle of “pathway analysis” (objective #2). This approach germinated a plethora of potential regulators of malignant phenomena, occurring at the level of the cancer invasion front. We focused on a small set of proteins, grouped together as “bone morphogenetic protein inhibitors” (BMPIs), which were then validated (objective #3). Following that, we developed a working model for the potential implication of the BMPIs in desmoplastic host-cell interfaces. We assumed that BMPIs could be overexpressed in the cancer invasion front to exert a BMP-dependent or even -independent dedifferentiation gradient in proximal cancer cells, and eventually induce mesenchymal phenotype (EMT program), and/or promote local invasion and tumor budding formation. To provide preliminary evidence in support of this working model, first, we performed co-expression studies (IHC), where the expression of various BMPIs was contextually correlated with invasion fronts undergoing an EMT processes and/or are characterized by tumor budding formation. Second, we lay groundwork in vitro assays, linking the BMP7/GREM1 axis to EMT and migration in various colon cancer cell lines (objectives #4-5).

7.1.2. Key Elements of the Experimental Pipeline

- We generated cocultures of SW480/SW620 colon cancer cell lines and normal colonic fibroblasts (18Co). We also indicated that these could recapitulate the recruitment of CAFs in the CRC invasion
fronts, since the 18Co fibroblasts exerted myofibroblastic phenotype transition in vitro, i.e. they overexpressed alpha-SMA, and the cocultures could secret elevated MMP2 and uPA.

• Using LC-MS/MS, we performed comprehensive proteomic analysis in the secretome of SW480/18Co and SW620/Co co-cultures, while the respective monocultures served as controls.

• Furthermore, by using the monoculture datasets as exclusion filters along with an in house-developed dendrogram for clustering secreted proteins, we proposed for the first time an in-depth proteomic signature of the exoproteome presented in the desmoplastic invasion front of CRC, which we termed “desmoplastic protein dataset” (DPD).

• Using pathway enrichment analysis (GSEA), we first organized the DPD signature into molecular themes of interest. An attractive, overrepresented theme that could be involved in local invasion and tumor budding formation was the “TGF-beta superfamily”, since these contextual signals have been long proposed to regulate EMT and to provide stem-like features in cancer cells.

• Subsequent protein-protein networks (Ingenuity/STRING) in the TGF-beta superfamily theme demonstrated that the theme was organized around certain molecular hubs, with most prominent ones, several inhibitors of the bone morphogenetic protein (BMP) pathway, including gremlin-1, follistatin, follistatin-like protein-3, high-temperature requirement-A3, and fibromodulin. Mainly due to availability of reagents, our focus was then turned towards GREM1, FST and HTRA3.

• The overrepresentation of GREM1, FST and HTRA3 was in vitro verified in both gene (RT-PCR) and protein (ELISA) expression levels, after reproducing relevant cocultures with the original proteomic platform. Moreover, biological relevance of BMPI overrepresentation was verified in a 30-patient cohort using immunohistochemistry, in which normal stromata had significantly lower expression of BMPIs compared to desmoplastic stromata.

• We demonstrated that HTRA3 expression was significantly associated with and could discriminate between microenvironments with or without tumor-budding in CRC.

• We demonstrated that GREM1 expression was significantly correlated with EMT-pertaining cancer cell cohorts invading the muscularis externae (intramuscular invasion fronts), as well as those forming nest-like configuration in the submucosa of CRC patients.

• We demonstrated that GREM1 could antagonize BMP7, and promote EMT in HT29 cells. This was shown both through morphological assessment of HT29 cell phenotype, as well as through the
regulation of prominent EMT markers in the gene expression level (RT-PCR), such as the shifting of the cadherin switch and nuclear localization of beta-catenin.

- We demonstrated that GREM1 could antagonize BMP7, and promote cell migration of both SW480 and SW620 cells, using an \textit{in-vitro} under-the-agarose cell migration assay.

\textbf{7.1.3. Working Model}

This thesis has provided some substantial evidence to support a working model of local invasion and tumor budding formation in advanced desmoplastic invasion fronts (Fig. 7.1.). This model is based on accumulating evidence which collectively support that the deployment of cell biological programs, such as epithelial-to-mesenchymal transition (EMT), at the level of the cancer invasion front, are necessary to effectively render stem-like properties to cancer cells and to eventually mediate local invasion and tumor-budding formation (Fig. 7.1.). The idea of the “context” is introduced to explain that a collective signature of microenvironmental signals should be present in these areas to be able to suppress the biological barriers of such metastatic behavior. The \textit{a posteriori} observation of this phenotype (i.e. local invasion and tumor budding) actually suggests that such context is indeed present and dynamically interacts and dominates over any potential metastasis-suppressor forces. A rate-limiting step and metastasis suppressor could be the release of signals from the proximal invasion fronts, which promote the maintenance and integrity of the epithelial phenotype, such as bone morphogenetic protein ligands (BMPs) (Fig. 7.1.). Indeed, it has been extensively shown that BMP pathway is a prominent barrier for the development and progression of various types of cancer, including the gastrointestinal cancers (i.e. colorectal cancer). As such, contextual signals released from reciprocal interactions with the cancer-associated fibroblasts (CAFs), i.e. paracrine signals of the desmoplastic microenvironment, could potentially disrupt the EMT-suppressing properties of the BMP pathway (Fig. 7.1.). Such signals have been identified through our proteomic approach in the current thesis, as selective or other type of BMP inhibitors/antagonists (i.e. BMPIs), which include GREM1, FST and HTRA3 (Fig 7.1.). Collectively, our efforts add some insights in this working model, supporting a notion that these BMPIs could be part of the many “ingredients” of a rather-unexplored \textit{context}, which could be sufficient and/or necessary for completing the metastatic propagation. Other ingredients that should be definitely prioritized as comprising incremental parts of this context are TGF-beta, Wnt, Hedgehog and Notch signals, among many others.\textsuperscript{583}

Recent evidence has connected the acquisition of cancer stem cell (CSC) (or cancer-initiating cell) traits, following an EMT transdifferentiation program.\textsuperscript{213,583} This thesis is not intended, at this point, to bridge the gap between CSCs and TBs, as being the exact same cancer subpopulation, despite the fact that this
notion has already been suggested in the incremental review by Hanahan and Weinberg, “the hallmarks of cancer” (2011)\(^2\). The origins of CSCs within a solid tumor have not yet been clarified, and may well vary from one tumor type to another. However, the induction of the EMT program in certain model systems is sufficient by itself to provide the so-called “stemness” in the target population, including the characteristics that define the stem-like phenotype, i.e. self-renewal capability and specific antigenic properties. Such notion may not only suggest that the EMT program can enable cancer cells to physically disseminate from primary tumors, but also confer self-renewal ability on them, which may be crucial for clonal expansion\(^{584, 585}\). Therefore, upon generalization, this connection raises a corollary model: the heterotypic signals that trigger an EMT, such as those released from an activated desmoplastic stroma (and in this thesis, we proposed BMPIs), could be sufficient and/or necessary for creating and maintaining cancer stem cell niches\(^{213, 586, 587}\).
FIGURE 7.1. Working model on the potential implication of BMP inhibitors, as contextual signals for mediating local invasion and tumor budding formation. The working model is based on accumulating evidence which collectively supports that the deployment of cell biological programs, such as epithelial-to-mesenchymal transition (EMT), at the level of the cancer invasion front, is necessary to effectively render stem-like properties to cancer cells, and eventually mediate local invasion and tumor-budding formation. A rate-limiting step for this process could be the release of signals from the proximal invasion fronts, which promote the maintenance and integrity of the epithelial phenotype, such as bone morphogenetic protein ligands (BMPs). Indeed, it has been extensively shown that BMP pathway is a prominent barrier for the development and progression of various types of cancer, including the gastrointestinal cancers (i.e. colorectal cancer). As such, contextual signals released from reciprocal interactions with cancer-associated fibroblasts (CAFs), i.e. paracrine signals derived from the desmoplastic microenvironment, could potentially disrupt the EMT-suppressing properties of the BMP pathway. Such signals have been identified through our proteomic approach in the current thesis, as selective or other type of BMP inhibitors/antagonists (i.e. BMPIs), which include GREM1, FST and HTRA3. However, BMP-independent mechanisms of EMT promotion cannot be excluded (demonstrated with question mark), as well as the importance of additional tumorigenic stromal responses in the tumor-host cell interface, such as the now well-established GREM1/VEGFR axis, shown to promote cancer-associated angiogenesis. Collectively, this working model supports a rationale that BMPIs could be one of the many “ingredients” of a rather unexplored context, which could be sufficient and/or necessary for completing the metastatic propagation. Other ingredients that could belong to this generic context are TGF-beta, Wnt, Hedgehog and Notch signals, among many others. Future work should prioritize the comprehension of this repertoire and provide a deep understanding of how these phenomena are regulated at the extracellular level.
7.2. FUTURE PERSPECTIVES

This thesis has laid fertile ground for answering very important questions in cancer pathobiology as to what could be the contextual signals that regulate important biological programs for localized invasion and metastasis, such as EMT, tumor budding formation and enhanced migration. However, due to the remarkable complexity of the desmoplastic interface, we continue to foresee that the implication of the BMP7-GREM1/HTRA3 axis in mediating EMT and tumor budding/local invasion is far from being entirely responsible for regulating the aforementioned processes. Therefore, this thesis is hoped to provide a framework for future research that attempts to elucidate these mechanisms, by proposing both biological and methodological innovations.

We anticipate that this thesis mostly serves as a proof-of-concept study, attempting to demonstrate the efficacy of mass spectrometry and proteomics in investigating tumor-host cell interactions. Undoubtedly, this thesis underscores the necessity that future efforts with proteomic technologies should focus on better experimental design to obtain pathophysiological relevance of CRC. In particular, cocultures that would cover all prominent CRC subtypes should also be developed, which include a wide series of MSS/MSI and CIMP+/- cell lines, and these secretomes/proteomes should be compared with each other, to generate more comprehensive and holistic molecular signatures of the desmoplastic invasion front. Moreover, the more-obsolete and problematic two-dimensional coculture model system should be replaced by the more relevant 3D-coculture systems that have documentedly showed that they can recapitulate pathophysiological aspects of cancers in a better way. Lastly, it should be mentioned that in vitro model systems should always be combined with other sources, such as tissue-based platforms. In the setting of the desmoplastic tumor-host cell interface, an attractive strategy would be the utilization of laser-capture microdissection (LCM) coupled to proteomic analysis, to explicitly isolate invasion fronts with or without tumor buds, and then uncover the specific microenvironmental milieu/context characterizing these diverse invasive behaviors through proteomics.

As already described, CRC is an extremely heterogeneous disease, where multiple levels of DNA instability have been shown to contribute to the development of distinct molecular pathways, whose independent progression is also reflected to a wide spectrum of polarized histological features and diverse phenotypes. In addition, these pathways might also be affected from a “field cancerization effect”, enrooted in prominent epigenetic differences across the various anatomical sites of the normal colon, where CRC may sporadically develop. Thus, we theorize that there is no consensus in trying to recapitulate the entire heterogeneity through simple cell line model systems, and as such, we strongly advocate that our proteomic approach using the SW480/SW620 model system simply reflects to a certain
subgroup of CRC patients. Given the (epi)genetic background of SW480/SW620 cells, we could-at best-support that our results could reflect in subgroups of CRC patients, exerting an MSS/CIMP- phenotype, despite the fact that epigenetic differences between SW480 and SW620 cell lines have been spotted. Following this logic, our approach raises certain concerns and limitations about all the preceding verification studies that have been conducted, especially about the immunohistochemical validation, where no stratification according to the (epi)genetic background was performed in the cohorts, either because such data were not directly available or because such stratification would create underpowered group sizes. This weakness has been quite expectedly reflected in many of our analyses; for instance, in the logistic regression model developed in chapter 5, we noted an adequate discriminatory power of HTRA3 in distinguishing tumor budding from invasion fronts without tumor budding; however, HTRA3 performance was relatively poor, possibly because this association is only meaningful for specific CRC subgroups and the true potential of HTRA3 in these subgroups was masked. Therefore, this thesis has raised a paradigm that future work should definitely include the utilization of larger patient cohorts, very well-characterized in terms of their (epi)genetic features, where signaling cascades and metastatic strategies would be explored in view of this comprehensive molecular profiling. It still remains to be confirmed, as to whether BMP antagonism is indeed an optimal strategy for regulating local invasion and tumor budding formation in MSS/CIMP- patients, as well as to which extent this signaling pathway could be meaningful to yet other subgroups of CRC patients.

A thorough investigation of the EMT-promoting properties of GREM1 (along with the rest BMPs identified) should be prioritized, to establish a more concrete causative association with local invasion and tumor budding formation. At this point, we simply provided proof of contextual correlation, which unfortunately does not necessarily support causation. To begin with, it will be essential to first evaluate the status of BMP signaling in CRC in general, since it is now known that BMP disruption may also occur in various levels downstream of the extracellular one (i.e. receptor, intracellular signaling, etc). As such, the exact contribution, if indeed any, of the extracellular BMP antagonism should be evaluated. Next, the downstream mediators should be investigated through traditional cell signaling and molecular biology techniques. Incremental questions, such as: (I) whether EMT is BMP-suppression-dependent or there are other BMP-independent mechanisms involved, (II) what type of signaling cascades mediate the EMT phenotype; canonical or non-canonical propagation? And finally, (III) what is the responsible interactome of this cascade, comprise only a few aspects that should be investigated, in the future.

It would also be incremental to examine the implication of the BMP7-GREM1/HTRA3 axis through loss-of-function studies in both the in vitro (i.e. specific shRNA-mediated knockdown in cell lines) and the in vivo (i.e. specific GREM1 and HTRA3 knockout animals) setting. Major efforts should encompass the
utilization of both GREM1-null and HTRA3-null mice to understand their roles in CRC carcinogenesis. Interestingly, GREM1-/− mice have been utilized in a number of developmental studies, but die shortly after birth, because of complete renal agenesis 302,304, an issue that renders this particular background rather inappropriate to study cancer in adult mice. However, it has been shown that heterozygous GREM1+/− mice are both viable and healthy, and can grow normally with no differences from wild type GREM1+/+ mice or vulnerability to any known diseases594. This animal model has been successfully used to study the implication of GREM1 in diabetic nephropathy in the past 594, and could be used in our context, as well. Technologies that would allow preferential disruption of target genes in fibroblasts (i.e. conditional knockouts) 595, would be another attractive strategy to circumvent GREM1-/− lethality and to investigate the role of this BMP antagonist in CRC progression. In regards with HTRA3, despite the fact that the expression pattern of this peptidase has been rather investigated in the context of ovary and maternal tissue development 318,596, HTRA3-null mice have not yet been developed.

Undoubtedly, an important question that needs to be addressed in the future is the elucidation of the exact mechanisms by which the expression of these BMP inhibitors is regulated in desmoplastic interfaces. Our immunohistochemical findings have consistently reported that both cancer cells and their associated fibroblasts could tentatively express GREM1, FST and HTRA3, and such observations are concordant with a speculative model that there is a very efficient on/off switch that mediates the process. As such, future research should definitely focus on the identification of the molecular switch that regulates the expression of these inhibitors/antagonists, especially if they prove to attain a tumor-promoting role, and targeted therapies against them are envisioned. To which direction, these research efforts should be headed, is quite difficult to think about, especially given the profound complexity and cross-signaling of the BMP with other intracellular pathways. Remarkably, the ability of cells to quickly turn on/off the expression of certain genes has been generally attributed to epigenetic mechanisms 597, and there is now accumulating evidence that many BMP antagonists, such as GREM1, are regulated through promoter methylation in many human cancers 531. Especially since our proposed model of BMP7/GREM1-mediated EMT in colorectal cancer cells of the invasion front could be actually interrelated with more generic models of cell reprogramming to pluripotency (i.e. “cancer-initiating cells” or “cancer stem cells”) 215,598, 599, their epigenetic landscape should be prioritized for such future investigations 597.

Furthermore, emerging evidence indicates alternative, BMP-independent mechanisms of promoting and sustaining tumorigenesis for many of these BMP antagonists 600. Of note, GREM1 has been shown to induce VEGFR-dependent angiogenesis in cancer 543,544,572, and GREM1 expression correlates with angiogenesis and progression-free survival in patients with pancreatic neuroendocrine tumors 601, as well as in patients with lung adenocarcinoma 527. The implication of BMP inhibitors in the acquisition of so
different cancer hallmarks provides a new dimension of thinking about them, and dictates future research to obtain insights in their pleiotropic effects. In particular, the investigation of GREM1/HTRA3 expression in relation to hyper-/hypovascularized areas in our patient cohorts, could be very easily performed in the near future, to assess whether they could be implicated in the shifting of the angiogenic switch in colorectal cancer, as well. Further exploration on the exact downstream signaling cascades could be chased afterwards, to propose more innovative and effective antiangiogenic therapies.

Finally, it should be kept in mind that the original proteomic analysis in this coculture model system has identified a 152-protein signature, the desmoplastic protein dataset (DPD), which could theoretically recapitulate additional contextual signals of the desmoplastic invasion front, but further investigations were prioritized for only three of them, GREM1, FST and HTRA3. The rationale behind selecting those three actually came from the pathway analysis described in chapter 3, whereby the TGF-beta superfamily aroused as a major candidate pathway for mediating the process of local invasion and tumor budding formation. However, it would be rather naïve to speculate that our selection criteria at that point were all the most optimal, in a sense that none of the remaining proteins of the DPD signature could harbor important implications in CRC progression. As such, future efforts should encompass more rationalized meta-mining in the acquired datasets, to further investigate and prioritize candidates with greater potential to mediate EMT and local invasion.

Taken together, we anticipate that with the information provided in this thesis, future research could expand to the direction of providing concrete, causative implication of the BMP7-GREM1/HTRA3 signaling cascade in CRC progression, as well as to examine for translational opportunities, if any, by scrutinizing this axis as potential biomarker of disease progression or even as therapeutic target.
CHAPTER 8

References


42. Egeblad, M.; Nakasone, E. S.; Werb, Z., Tumors as organs: complex tissues that interface with the entire organism. *Dev Cell* 2010, 18, (6), 884-901.


63. Xing, F.; Saidou, J.; Watabe, K., Cancer associated fibroblasts (CAFs) in tumor microenvironment. *Front Biosci (Landmark Ed)* 2009, 15, 166-79.


169


Zhang, H.; Bradley, A., Mice deficient for BMP2 are nonviable and have defects in amnion/chorion and cardiac development. *Development* 1996, 122, (10), 2977-86.


424. Lao, V. V.; Grady, W. M., Epigenetics and colorectal cancer. *Nat Rev Gastroenterol Hepatol* 2013, 8, (12), 686-700.


491. Glaros, T.; Larsen, M.; Li, L., Macrophages and fibroblasts during inflammation, tissue damage and organ injury. Front Biosci (Landmark Ed) 2009, 14, 3988-93.


CHAPTER 9

Appendix
Table A. The Desmoplastic Protein Dataset (list of proteins in alphabetical order)

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACP1 Isoform 2</td>
<td>Low molecular weight phosphotyrosine protein phosphatase</td>
</tr>
<tr>
<td>ADAMTS12</td>
<td>ADAM metallopeptidase with thrombospondin type 1 motif, 12 preproprotein</td>
</tr>
<tr>
<td>AEBP1 Isoform 1</td>
<td>Adipocyte enhancer-binding protein 1</td>
</tr>
<tr>
<td>APOB B-100</td>
<td>Apolipoprotein B-100</td>
</tr>
<tr>
<td>APOC2 C-II</td>
<td>Apolipoprotein C-II</td>
</tr>
<tr>
<td>ARF6</td>
<td>ADP-ribosylation factor 6</td>
</tr>
<tr>
<td>ARSA</td>
<td>Arylsulfatase A</td>
</tr>
<tr>
<td>ASPH</td>
<td>Aspartyl/asparaginyl beta-hydroxylase</td>
</tr>
<tr>
<td>ATP1A1 Isoform Long</td>
<td>Sodium/potassium-transporting ATPase subunit alpha-1</td>
</tr>
<tr>
<td>ATP1B3</td>
<td>Sodium/potassium-transporting ATPase subunit beta-3</td>
</tr>
<tr>
<td>ATP5J ATP synthase Coupling Factor 6, mitochondrial</td>
<td></td>
</tr>
<tr>
<td>AZGP1</td>
<td>Zinc-alpha-2-glycoprotein</td>
</tr>
<tr>
<td>B4GALT4 Beta-1,4-galactosyltransferase 4</td>
<td></td>
</tr>
<tr>
<td>BAX Isoform Alpha</td>
<td>Apoptosis regulator BAX</td>
</tr>
<tr>
<td>BCAP31 B-cell receptor-associated protein 31</td>
<td></td>
</tr>
<tr>
<td>C10orf116 Adipose most abundant gene transcript 2</td>
<td>protein</td>
</tr>
<tr>
<td>C14orf156 SRA stem-loop-interacting RNA-binding</td>
<td>mitochondrial</td>
</tr>
<tr>
<td>C1QTNF5 Complement C1q tumor necrosis factor-related protein 5</td>
<td></td>
</tr>
<tr>
<td>C4A Putative uncharacterized protein C4A</td>
<td></td>
</tr>
<tr>
<td>C4B complement component 4B preproprotein</td>
<td></td>
</tr>
<tr>
<td>CCBE1 Isoform 1</td>
<td>Collagen and calcium-binding EGF domain-containing protein 1</td>
</tr>
<tr>
<td>CCDC72 Coiled-coil domain-containing protein 72</td>
<td></td>
</tr>
<tr>
<td>CD81</td>
<td>CD81 antigen</td>
</tr>
<tr>
<td>CD99 Isoform 1</td>
<td>CD99 antigen</td>
</tr>
<tr>
<td>CDH11 Isoform 2</td>
<td>Cadherin-11</td>
</tr>
<tr>
<td>CDH2 Cadherin-2</td>
<td></td>
</tr>
<tr>
<td>CDH6 Isoform 1</td>
<td>Cadherin-6</td>
</tr>
<tr>
<td>CFD Complement factor D preproprotein</td>
<td></td>
</tr>
<tr>
<td>CFH Isoform 1</td>
<td>Complement factor H</td>
</tr>
<tr>
<td>CLSTN1 Isoform 1</td>
<td>Calsyntenin-1 (Fragment)</td>
</tr>
<tr>
<td>CLTA Isoform Brain</td>
<td>Clathrin light chain A</td>
</tr>
<tr>
<td>CNBP I Isoform 1</td>
<td>Cellular nucleic acid-binding protein</td>
</tr>
<tr>
<td>COL12A1 Isoform 4</td>
<td>Collagen alpha-1(XII) chain</td>
</tr>
<tr>
<td>COL16A1 Isoform 1</td>
<td>Collagen alpha-1(XVI) chain</td>
</tr>
<tr>
<td>COL4A1 Isoform 1</td>
<td>Collagen alpha-1(IV) chain</td>
</tr>
<tr>
<td>COL4A1 Putative uncharacterized protein COL4A1</td>
<td></td>
</tr>
<tr>
<td>COL7A1 Isoform 1</td>
<td>Collagen alpha-1(VII) chain</td>
</tr>
<tr>
<td>COLEC10 Collectin-10</td>
<td></td>
</tr>
<tr>
<td>COLEC12 Isoform 1</td>
<td>Collectin-12</td>
</tr>
<tr>
<td>CPA4 Carboxypeptidase A4</td>
<td></td>
</tr>
<tr>
<td>CRISPLD2 Isoform 1</td>
<td>Cysteine-rich secretory protein LCCL domain-containing 2</td>
</tr>
<tr>
<td>CRTAP Cartilage-associated protein</td>
<td></td>
</tr>
<tr>
<td>Gene Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>CSF2</td>
<td>Granulocyte-macrophage colony-stimulating factor</td>
</tr>
<tr>
<td>CST7</td>
<td>Cystatin-F precursor</td>
</tr>
<tr>
<td>CTHRC1</td>
<td>Isoform 1 of Collagen triple helix repeat-containing protein 1</td>
</tr>
<tr>
<td>CTSC</td>
<td>Dipeptidyl peptidase 1</td>
</tr>
<tr>
<td>CXCL10</td>
<td>C-X-C motif chemokine 10</td>
</tr>
<tr>
<td>DKK3</td>
<td>Dickkopf-related protein 3</td>
</tr>
<tr>
<td>DNAJC3</td>
<td>Isoform 1 of DnaJ homolog subfamily C member 3</td>
</tr>
<tr>
<td>DPEP1</td>
<td>Dipeptidase 1</td>
</tr>
<tr>
<td>DSC2</td>
<td>Isoform 2A of Desmocollin-2</td>
</tr>
<tr>
<td>EIF3F</td>
<td>cDNA FLJ36192 fis, clone TESTI2027450, highly similar to Eukaryotic translation initiation factor 3 subunit 5</td>
</tr>
<tr>
<td>ENPP2</td>
<td>Isoform Long of Endoglin</td>
</tr>
<tr>
<td>ERP44</td>
<td>Endoplasmic reticulum resident protein 44</td>
</tr>
<tr>
<td>EXT2</td>
<td>Isoform 1 of Exostosin-2</td>
</tr>
<tr>
<td>EXTL2</td>
<td>Exostoses (Multiple)-like 2</td>
</tr>
<tr>
<td>FAM20C</td>
<td>family with sequence similarity 20, member C</td>
</tr>
<tr>
<td>FAT1</td>
<td>507 kDa protein</td>
</tr>
<tr>
<td>FBLN2</td>
<td>cDNA FLJ56912, highly similar to Fibulin-2</td>
</tr>
<tr>
<td>FDPS</td>
<td>Farnesyl pyrophosphate synthase</td>
</tr>
<tr>
<td>FIBIN</td>
<td>Fin bud initiation factor homolog</td>
</tr>
<tr>
<td>FKB10</td>
<td>Peptidyl-prolyl cis-trans isomerase FKB10</td>
</tr>
<tr>
<td>FMOD</td>
<td>Fibromodulin</td>
</tr>
<tr>
<td>FST</td>
<td>Isoform 1 of Follistatin</td>
</tr>
<tr>
<td>FSTL3</td>
<td>Isoform 1 of Follistatin-related protein 3</td>
</tr>
<tr>
<td>GALNT10</td>
<td>Isoform 1 of Polypeptide N-acetylgalactosaminyltransferase 10</td>
</tr>
<tr>
<td>GALNT5</td>
<td>Polypeptide N-acetylgalactosaminyltransferase 5</td>
</tr>
<tr>
<td>GDF6</td>
<td>Growth/differentiation factor 6</td>
</tr>
<tr>
<td>GFRA1</td>
<td>Isoform 2 of GDNF family receptor alpha-1</td>
</tr>
<tr>
<td>GREM1</td>
<td>Isoform 1 of Gremlin-1</td>
</tr>
<tr>
<td>HDAC2</td>
<td>Isoform 2 of Histone deacetylase 2</td>
</tr>
<tr>
<td>HEXA</td>
<td>Beta-hexosaminidase subunit alpha</td>
</tr>
<tr>
<td>HGFR</td>
<td>Isoform 1 of Hepatocyte growth factor</td>
</tr>
<tr>
<td>HLA-A</td>
<td>HLA class I histocompatibility antigen, A-34 alpha chain</td>
</tr>
<tr>
<td>HLA-A</td>
<td>HLA class I histocompatibility antigen, A-69 alpha chain</td>
</tr>
<tr>
<td>HLA-C</td>
<td>HLA class I histocompatibility antigen, Cw-7 alpha chain</td>
</tr>
<tr>
<td>HSD17B10</td>
<td>Isoform 1 of 3-hydroxyacyl-CoA dehydrogenase type-2</td>
</tr>
<tr>
<td>HTRA3</td>
<td>Isoform 1 of Probable serine protease HTRA3</td>
</tr>
<tr>
<td>IDUA</td>
<td>Alpha-L-iduronidase</td>
</tr>
<tr>
<td>IGFR</td>
<td>Cation-independent mannose-6-phosphate receptor</td>
</tr>
<tr>
<td>IGFBP2</td>
<td>insulin-like growth factor binding protein 2, 36kDa precursor</td>
</tr>
<tr>
<td>IL6ST</td>
<td>Isoform 1 of Interleukin-6 receptor subunit beta</td>
</tr>
<tr>
<td>INHBA</td>
<td>Inhibin beta A chain</td>
</tr>
<tr>
<td>IPO9</td>
<td>Importin-9</td>
</tr>
<tr>
<td>ITGA2</td>
<td>Integrin alpha-2</td>
</tr>
<tr>
<td>ITGA5</td>
<td>Integrin alpha-5</td>
</tr>
</tbody>
</table>
ITGB1 Isoform Beta-1A of Integrin beta-1
ITIH3 Isoform 1 of Inter-alpha-trypsin inhibitor heavy chain H3
LAMA1 Laminin subunit alpha-1
LAMA3 Laminin alpha-3 chain variant 1
LAMB1 200 kDa protein
LAYN Isoform 1 of Layilin
LEPRE1 Isoform 3 of Prolyl 3-hydroxylase 1
LOC100293101;LOC100288413;LOC100290734 Uncharacterized protein LP9056
LOC284297 Scavenger receptor cysteine-rich domain-containing protein LOC284297
LOC728825 Putative uncharacterized protein ENSP0000391266
LOX Protein-lysine 6-oxidase
LTBP1 Isoform Short of Latent-transforming growth factor beta-binding protein 1
LTBP1 latent transforming growth factor beta binding protein 1 isoform LTBP-1S
MAMDC2 Isoform 1 of MAM domain-containing protein 2
MMP11 Stromelysin-3
MRC2 C-type mannose receptor 2
MT1L;MT1JP;MT1G;MT1E Isoform 1 of Metallothionein-1G
MXRA8 Isoform 2 of Matrix-remodeling-associated protein 8
NEU1 Sialidase
NFASC Isoform 7 of Neurofascin
NID2 Nidogen-2
PAMR1 Isoform 2 of Inactive serine protease RAMP
PAPPA Pappalysin-1
PCDHGC3 Isoform 1 of Protocadherin gamma-C3
PCMT1 protein-L-isoaspartate (D-aspartate) O-methyltransferase 1
PHB Prohibitin
PIK3P1 Isoform 1 of Phosphoinositide-3-kinase-interacting protein 1
PLA2G15 Group XV phospholipase A2
PLAU Urokinase-type plasminogen activator
PLXNB2 Plexin-B2
POSTN Periostin, osteoblast specific factor
PP1F Peptidyl-prolyl cis-trans isomerase F, mitochondrial
PRNP Major prion protein
PROCR Endothelial protein C receptor precursor
PSG1 Isoform 2 of Pregnancy-specific beta-1-glycoprotein 1
PSMC2 26S protease regulatory subunit 7
PSMD14 26S proteasome non-ATPase regulatory subunit 14
QPCT Isoform 1 of Glutaminyl-peptide cyclotransferase
RPL21;RPL21P19 60S ribosomal protein L21
RPL21P19;RPL21 60S ribosomal protein L21
RPS11 40S ribosomal protein S11
RPS21;LOC100291837 40S ribosomal protein S21
RPS27A;UBB;UBC ubiquitin and ribosomal protein S27a precursor
SDCBP Isoform 1 of Syntenin-1
SECTM1 Secreted and transmembrane protein 1
SELM Selenoprotein M
SEMA7A Semaphorin-7A
SIAE Isoform 1 of Sialate O-acetylesterase
SIL1 Nucleotide exchange factor SIL1
SLC25A5 ADP/ATP translocase 2
SRPX2 Sushi repeat-containing protein SRPX2
SSC5D Scavenger receptor cysteine-rich domain-containing protein LOC284297
STC1 Stanniocalcin-1
TBCB Tubulin-folding cofactor B
TFF3 14 kDa protein
TGFB2 Isoform B of Transforming growth factor beta-2
TGFBR3 transforming growth factor, beta receptor III
TGOLN2 Isoform TGN51 of Trans-Golgi network integral membrane protein 2
THSD4 Isoform 1 of Thrombospondin type-1 domain-containing protein 4
TOMM34 Mitochondrial import receptor subunit TOM34
TPD52L2 Isoform 2 of Tumor protein D54
VAPA Isoform 1 of Vesicle-associated membrane protein-associated protein A
VBP1 von Hippel-Lindau binding protein 1, isoform CRA_b
VCAN Isoform V0 of Versican core protein
VCAN Isoform V1 of Versican core protein
Table B. Patient demographic data (validation cohort; N=38), used for correction in the logistic regression model

<table>
<thead>
<tr>
<th>Case</th>
<th>Gender</th>
<th>Age</th>
<th>T-category</th>
<th>LN-group</th>
<th>Context</th>
<th>Crohns-Like Lesions</th>
<th>Tumor Location</th>
<th>Side</th>
<th>HTRA3 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>76</td>
<td>T4</td>
<td>&gt;15</td>
<td>TB</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>sigmoid</td>
<td>Left 3</td>
</tr>
<tr>
<td>2</td>
<td>Male</td>
<td>82</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>recto-sigmoid</td>
<td>Left 1.7</td>
</tr>
<tr>
<td>3</td>
<td>Male</td>
<td>61</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Marked</td>
<td>Absent</td>
<td>cecum</td>
<td>Right 1.4</td>
</tr>
<tr>
<td>4</td>
<td>Female</td>
<td>70</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Present</td>
<td>transverse</td>
<td>Right 1.7</td>
</tr>
<tr>
<td>5</td>
<td>Male</td>
<td>71</td>
<td>T4</td>
<td>&gt;15</td>
<td>TB</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>cecum</td>
<td>Right 2.5</td>
</tr>
<tr>
<td>6</td>
<td>Male</td>
<td>67</td>
<td>T3</td>
<td>&gt;15</td>
<td>TB</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>ascending</td>
<td>Right 3.2</td>
</tr>
<tr>
<td>7</td>
<td>Female</td>
<td>63</td>
<td>T4</td>
<td>&lt;12</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>cecum</td>
<td>Right 2.7</td>
</tr>
<tr>
<td>8</td>
<td>Male</td>
<td>59</td>
<td>T3</td>
<td>12-15</td>
<td>TB</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>cecum</td>
<td>Right 2.2</td>
</tr>
<tr>
<td>9</td>
<td>Male</td>
<td>66</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>sigmoid</td>
<td>Left 1.5</td>
</tr>
<tr>
<td>10</td>
<td>Male</td>
<td>80</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>sigmoid</td>
<td>Left 0</td>
</tr>
<tr>
<td>11</td>
<td>Female</td>
<td>80</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Marked</td>
<td>Absent</td>
<td>cecum</td>
<td>Right 1.8</td>
</tr>
<tr>
<td>12</td>
<td>Male</td>
<td>86</td>
<td>T4</td>
<td>&gt;15</td>
<td>TB</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>cecum</td>
<td>Right 2.4</td>
</tr>
<tr>
<td>13</td>
<td>Male</td>
<td>51</td>
<td>T4</td>
<td>&gt;15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>sigmoid</td>
<td>Left 0.8</td>
</tr>
<tr>
<td>14</td>
<td>Male</td>
<td>78</td>
<td>T3</td>
<td>&gt;15</td>
<td>TB</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>rectosigmoid</td>
<td>Left 3.2</td>
</tr>
<tr>
<td>15</td>
<td>Male</td>
<td>74</td>
<td>T4</td>
<td>&gt;15</td>
<td>TB</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>cecum</td>
<td>Right 2.1</td>
</tr>
<tr>
<td>16</td>
<td>Male</td>
<td>63</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>descending</td>
<td>Left 0.8</td>
</tr>
<tr>
<td>17</td>
<td>Male</td>
<td>79</td>
<td>T4</td>
<td>12-15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>sigmoid</td>
<td>Left 1</td>
</tr>
<tr>
<td>18</td>
<td>Female</td>
<td>89</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Marked</td>
<td>Absent</td>
<td>ascending</td>
<td>Right 1</td>
</tr>
<tr>
<td>19</td>
<td>Male</td>
<td>73</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>sigmoid</td>
<td>Left 1.8</td>
</tr>
<tr>
<td>20</td>
<td>Male</td>
<td>61</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>cecum</td>
<td>Right 3.2</td>
</tr>
<tr>
<td>21</td>
<td>Female</td>
<td>87</td>
<td>T3</td>
<td>&gt;15</td>
<td>TB</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>sigmoid</td>
<td>Left 2.9</td>
</tr>
<tr>
<td>22</td>
<td>Female</td>
<td>77</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Marked</td>
<td>Present</td>
<td>sigmoid</td>
<td>Left 1</td>
</tr>
<tr>
<td>23</td>
<td>Female</td>
<td>56</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>sigmoid</td>
<td>Left 0.6</td>
</tr>
<tr>
<td>24</td>
<td>Female</td>
<td>34</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Marked</td>
<td>Absent</td>
<td>ascending</td>
<td>Right 0.8</td>
</tr>
<tr>
<td>25</td>
<td>Female</td>
<td>58</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>cecum</td>
<td>Right 2.6</td>
</tr>
<tr>
<td>26</td>
<td>Male</td>
<td>79</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>sigmoid</td>
<td>Left 1.1</td>
</tr>
<tr>
<td>27</td>
<td>Female</td>
<td>72</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>rectosigmoid</td>
<td>Left 1.3</td>
</tr>
<tr>
<td>28</td>
<td>Female</td>
<td>60</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>cecum</td>
<td>Right 2.1</td>
</tr>
<tr>
<td>29</td>
<td>Female</td>
<td>40</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Marked</td>
<td>Present</td>
<td>descending</td>
<td>Left 1.2</td>
</tr>
<tr>
<td>30</td>
<td>Female</td>
<td>64</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Marked</td>
<td>Absent</td>
<td>cecum</td>
<td>Right 1.5</td>
</tr>
<tr>
<td>31</td>
<td>Female</td>
<td>56</td>
<td>T3</td>
<td>&gt;15</td>
<td>TB</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>transverse</td>
<td>Right 4.8</td>
</tr>
<tr>
<td>32</td>
<td>Female</td>
<td>75</td>
<td>T3</td>
<td>12-15</td>
<td>TB</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>sigmoid</td>
<td>Left 4.5</td>
</tr>
<tr>
<td>33</td>
<td>Male</td>
<td>68</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>splenic flexure</td>
<td>Right 1.8</td>
</tr>
<tr>
<td>34</td>
<td>Female</td>
<td>64</td>
<td>T3</td>
<td>&gt;15</td>
<td>IF</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>ascending</td>
<td>Right 2.1</td>
</tr>
<tr>
<td>35</td>
<td>Female</td>
<td>53</td>
<td>T3</td>
<td>&gt;15</td>
<td>TB</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>ascending</td>
<td>Right 3.6</td>
</tr>
<tr>
<td>36</td>
<td>Male</td>
<td>77</td>
<td>T3</td>
<td>&gt;15</td>
<td>TB</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>ascending</td>
<td>Right 2.1</td>
</tr>
<tr>
<td>37</td>
<td>Male</td>
<td>61</td>
<td>T3</td>
<td>&gt;15</td>
<td>TB</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>sigmoid</td>
<td>Left 2</td>
</tr>
<tr>
<td>38</td>
<td>Male</td>
<td>56</td>
<td>T4</td>
<td>&gt;15</td>
<td>TB</td>
<td>Absent/Mild</td>
<td>Absent</td>
<td>rectum</td>
<td>Left 3.4</td>
</tr>
</tbody>
</table>
FIGURE A. Enriched GO terms for the genes found in the three TGF-beta-superfamily-associated clusters.
FIGURE B. Categorization of proteins belonging to the TGF-beta superfamily of proteins into group I (Ligands/Receptors) and group II (Regulators/Inhibitors).
FIGURE C: In-built negative controls. Left figure, magnification X200, FST staining, asterisk demonstrates lack of immunoreactivity in smooth muscle cells; middle figure, magnification X200, HTRA3 staining, lack of immunoreactivity in smooth muscle cells; right figure, magnification X100, GREM1 staining, asterisk demonstrates lack of immunoreactivity in immune cells, note also GREM1 immunoreactivity staining myofibroblastic niches around numerous healthy colonic crypts (positive control).
FIGURE D. Pattern of BMP inhibitor expression across four quartiles of increasing BMP antagonism. Values are shown as mean expression scores with standard errors. Red-dashed lines: mean expression score of tBMP1 in the corresponding quartile; black arrows: peaks corresponding to the conserved eGREM1, sHTRA3 and sFST determinants; red arrow: newly-appearing peak in the Q4 quartile corresponding to the eHTRA3 determinant.
FIGURE E. Alkaline phosphatase activity, following BMP7 treatment in various colon cancer cell lines. Note the prominent increase in ALP activity following BMP7 treatment in all cancer cell lines but the RKO. Treatments were performed in triplicates, and each data point represented the mean OD405 of the three replicates with the respective standard deviation.
FIGURE F. Optimization of the under-the-agarose cell migration assay. The punching of the central and peripheral cavities was performed as explained in the manuscript. In this experiment, all four peripheral wells were filled with DMEM supplemented with 10% FBS to generate a chemotactic gradient. No additional motogens were utilized. Interestingly, approximately 25% of the total cell population migrated towards each individual quadrant. There was no statistically significant difference between the intensities observed in all four quadrants (ANOVA; p>0.05), suggesting that potential differences in the actual experimental setups could be attributed to the addition of motogens in certain wells.