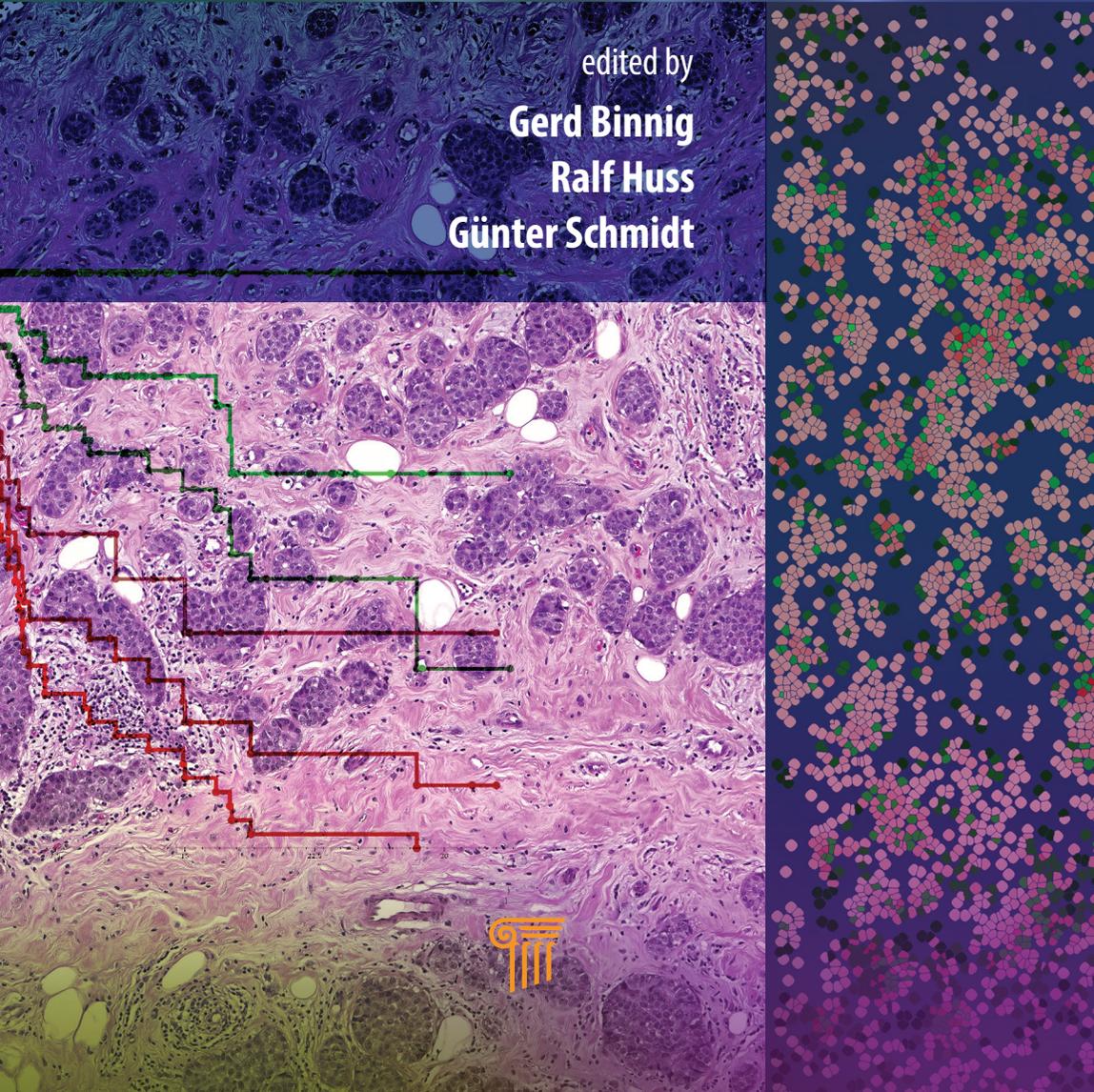


Jenny Stanford Series on Next-Generation Medicine Vol. 1

# Tissue Phenomics

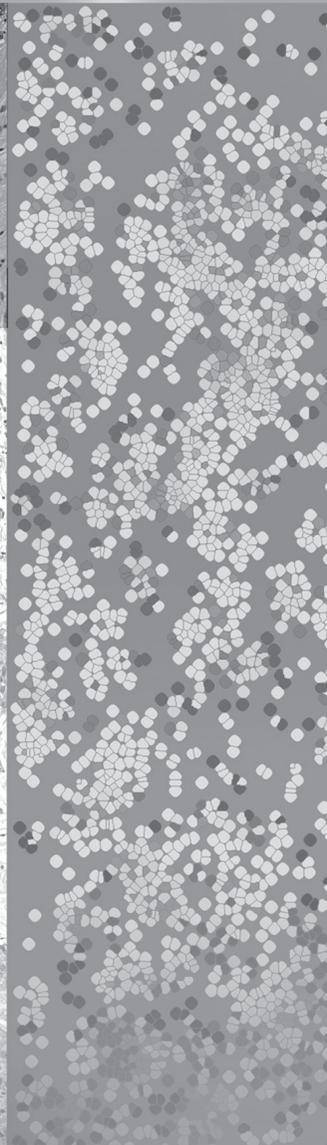
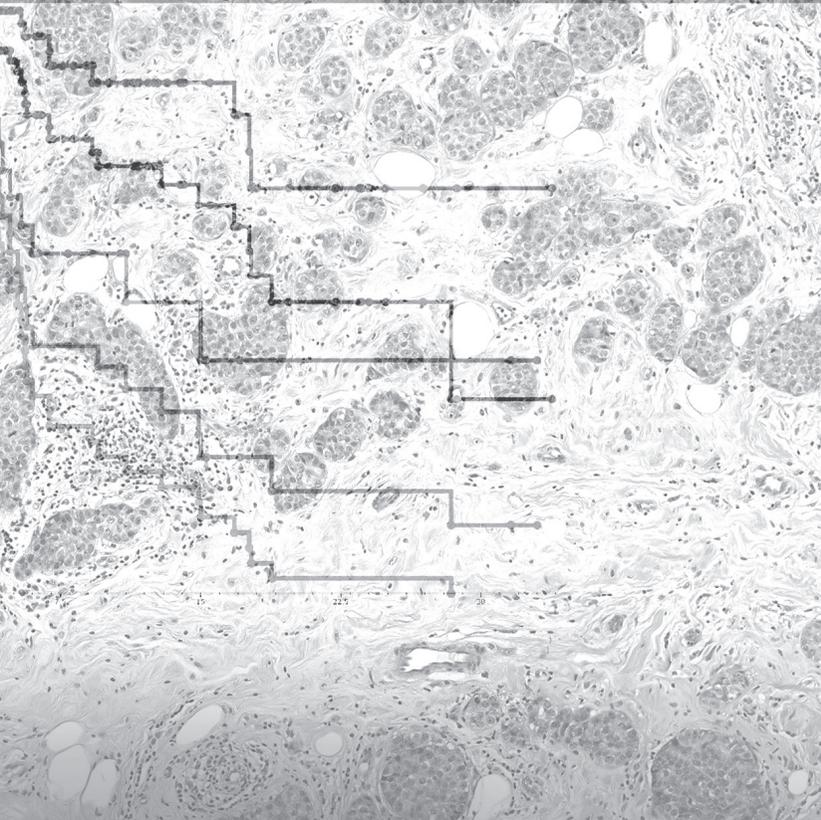
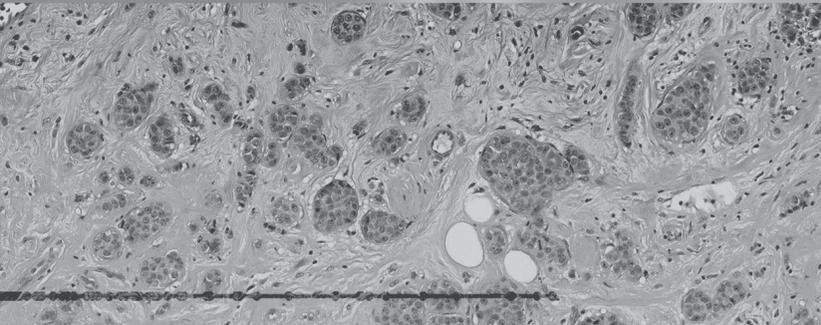
Profiling Cancer Patients for  
Treatment Decisions

edited by  
**Gerd Binnig**  
**Ralf Huss**  
**Günter Schmidt**





# Tissue Phenomics





Jenny Stanford Series on Next-Generation Medicine Vol. 1

# Tissue Phenomics

Profiling Cancer Patients for  
Treatment Decisions

edited by

**Gerd Binnig**

**Ralf Huss**

**Günter Schmidt**



JENNY STANFORD  
PUBLISHING

*Published by*

Jenny Stanford Publishing Pte. Ltd.  
Level 34, Centennial Tower  
3 Temasek Avenue  
Singapore 039190

Email: [editorial@jennystanford.com](mailto:editorial@jennystanford.com)

Web: [www.jennystanford.com](http://www.jennystanford.com)

**British Library Cataloguing-in-Publication Data**

A catalogue record for this book is available from the British Library.

**Tissue Phenomics: Profiling Cancer Patients for  
Treatment Decisions**

Copyright © 2019 by Jenny Stanford Publishing Pte. Ltd.

*All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the publisher.*

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 978-981-4774-88-8 (Hardcover)

ISBN 978-1-351-13427-9 (eBook)

*Morphological and functional complexity of animal and human tissue is one of the most fascinating and simultaneously challenging topics in tissue-based science. The diverse organizational units of a normal liver, lung, or kidney, such as organ-specific cells, nerves, connective tissues, and different types of vessels, show so many variants that a systematic and comprehensive analysis by human eyes is not possible. In addition, time-related modifications and spatial distribution of the components as well as disease-related variants produce an even higher level of complexity, often termed hyper-complexity.*

*Today, the only way to find a solution for reliable and reproducible analyses of various tissues is based on multiplex digital systems that produce tissue-related big data. By applying suitable algorithms, these data can be sorted and used to answer different diagnostic, prognostic, predictive, and scientific questions. The concept of tissue phenomics is currently the most promising approach to answer many burning questions of cancer and other diseases.*

**—Prof. Dr. med. Dr. h. c. Manfred Dietel**  
Charité, Berlin, Germany



# Contents

<i>Foreword</i>	xiii
<i>Acknowledgments</i>	xvii
<b>1. Introduction to Tissue Phenomics</b>	<b>1</b>
<i>Ralf Huss, Gerd Binnig, Günter Schmidt, Martin Baatz, and Johannes Zimmermann</i>	
1.1 Motivation	1
1.2 Tissue Phenomics and Medicine 4.0	3
1.3 Phenomena in Cancer Immunology	5
1.4 Future of Tissue Phenomics	6
1.5 About the Book	7
<b>2. Image Analysis for Tissue Phenomics</b>	<b>9</b>
<i>Johannes Zimmermann, Keith E. Steele, Brian Laffin, René Korn, Jan Lesniak, Tobias Wiestler, and Martin Baatz</i>	
2.1 Introduction	10
2.2 Experimental Design for Image Analysis Studies	12
2.3 Input Data and Test Data	14
2.4 Image Analysis	15
2.5 Quality Control Procedures	25
2.6 Results and Output	30
2.7 Discussion and Summary	31
<b>3. Context-Driven Image Analysis: Cognition Network Language</b>	<b>35</b>
<i>Gerd Binnig</i>	
3.1 Motivation and Reasoning	36
3.2 History and Philosophy	40
3.3 Technology	43
3.3.1 Class Hierarchy	44
3.3.2 Process Hierarchy and Context Navigation	44

	3.3.2.1	Processes and the domain concept	44
	3.3.2.2	Context navigation	45
	3.3.2.3	Object-based processing	46
	3.3.2.4	Maps	48
	3.3.2.5	Object variables	49
3.4		Example of Context-Driven Analysis	50
	3.4.1	H&E Analysis Problems	53
	3.4.2	Concrete H&E Image Analysis Example	54
	3.4.2.1	Color channel analysis as a starting point	54
	3.4.2.2	Isolated nuclei: first context objects	55
	3.4.2.3	Some remaining large nuclei by splitting and <i>inside-out</i> procedure	60
	3.4.2.4	Nuclei of immune cells	61
	3.4.2.5	Template match for heterogeneous nuclei on 10× and 20×	63
3.5		Conclusion	65
<b>4.</b>		<b>Machine Learning: A Data-Driven Approach to Image Analysis</b>	<b>69</b>
		<i>Nicolas Brieu, Maximilian Baust, Nathalie Harder, Katharina Nekolla, Armin Meier, and Günter Schmidt</i>	
4.1		Introduction: From Knowledge-Driven to Data-Driven Systems	70
4.2		Basics of Machine Learning	71
	4.2.1	Supervised Learning	72
	4.2.2	Classification and Regression Problems	72
	4.2.3	Data Organization	73
4.3		Random Forests for Feature Learning	77
	4.3.1	Decision Trees	77
	4.3.1.1	Definition	77
	4.3.1.2	Decision function	78
	4.3.1.3	Extremely randomized trees	78
	4.3.1.4	Training objective function	79
	4.3.2	Random Forests Ensemble (Bagging)	79
	4.3.3	Model Parameters	80

4.3.4	Application Generic Visual Context Features	81
4.3.4.1	Haar-like features	81
4.3.4.2	Gabor features	82
4.3.5	Application to the Analysis of Digital Pathology Images	82
4.3.5.1	On-the-fly learning of slide-specific random forests models	83
4.3.5.2	Area-guided distance-based detection of cell centers	84
4.4	Deep Convolutional Neural Networks	86
4.4.1	History: From Perceptrons and the XOR Problem to Deep Networks	87
4.4.2	Building Blocks	89
4.4.2.1	Convolutional layers	89
4.4.2.2	Convolutional neural networks	90
4.4.2.3	Loss functions	91
4.4.2.4	Activation functions	91
4.4.2.5	Pooling layers	92
4.4.2.6	Dropout layers	93
4.4.3	Application Examples	94
4.5	Discussion and Conclusion	97
4.5.1	Model Complexity and Data Availability	97
4.5.2	Knowledge and Data Teaming	97
4.5.3	Machine Learning Teaming	98
4.5.4	Conclusion	98
<b>5.</b>	<b>Image-Based Data Mining</b>	<b>101</b>
	<i>Ralf Schönmeier, Arno Schäpe, and Günter Schmidt</i>	
5.1	Introduction	102
5.2	Generating High-Level Features for Patient Diagnosis	104
5.2.1	Quantification of Regions of Interest	104
5.2.2	Heatmaps	105
5.2.2.1	Contents of heatmaps	106
5.2.2.2	Visualization of heatmaps	107
5.2.2.3	Heatmaps with multiplexed data	108

5.2.2.4	Objects in heatmaps	110
5.2.3	Algebraic Feature Composition	111
5.2.4	Aggregation of Multiple Tissue Pieces to Single Patient Descriptors	113
5.2.5	Integration of Clinical Data and Other Omics Data	113
5.3	Performance Metrics	115
5.4	Feature Selection Methods	117
5.4.1	Unsupervised Methods	117
5.4.2	Hypothesis-Driven Methods	118
5.4.3	Univariate, Data-Driven Feature Selection	120
5.5	Tissue Phenomics Loop	121
5.6	Tissue Phenomics Software	123
5.6.1	Image Analysis and Data Integration	123
5.6.2	General Architecture	123
5.6.3	Image Mining Software	124
5.6.4	Data and Workflow Management and Collaboration	125
5.7	Discussion and Outlook	127
<b>6.</b>	<b>Bioinformatics</b>	<b>131</b>
	<i>Sriram Sridhar, Brandon W. Higgs, and Sonja Althammer</i>	
6.1	Molecular Technologies: Past to Present	132
6.2	Genomics and Tissue Phenomics	134
6.2.1	Genomics Data Sources	134
6.2.2	The Art of Image Mining	138
6.2.2.1	Cell-to-cell distances	139
6.2.2.2	Quantifying cell populations in histological regions	140
6.2.2.3	Tissue phenome and survival	142
6.2.3	Power of Integrative Approaches	142
6.3	Analytical Approaches for Image and Genomics Data	144
6.3.1	Data Handling Requires Similar Methods	144
6.3.2	Enhancing Confidence in a Discovery	145
6.4	Examples of Genomics or IHC Biomarkers in Clinical Practice	147
6.4.1	Biomarker Background	147

6.4.2	Genomics and IHC to Guide Prognosis or Diagnosis	148
6.4.3	Patient Stratification: Genomics and IHC to Identify Patient Subsets for Treatment	150
<b>7.</b>	<b>Applications of Tissue Phenomics</b>	<b>157</b>
	<i>Johannes Zimmermann, Nathalie Harder, and Brian Laffin</i>	
7.1	Introduction	158
7.2	Hypothesis-Driven Approaches	159
7.2.1	TME as a Battlefield: CD8 and PD-L1 Densities Facilitate Patient Stratification for Durvalumab Therapy in Non-Small Cell Lung Cancer	159
7.2.2	Gland Morphology and TAM Distribution Patterns Outperform Gleason Score in Prostate Cancer	159
7.2.3	Novel Spatial Features Improve Staging in Colorectal Cancer	160
7.2.4	Immunoscore Is a Novel Predictor of Patient Survival in Colorectal Cancer	162
7.2.5	Analysis of Spatial Interaction Patterns Improves Conventional Cell Density Scores in Breast Cancer	163
7.2.6	Immune Cell Infiltration is Prognostic in Breast Cancer	165
7.2.7	The Immune Landscape Structure Directly Corresponds with Clinical Outcome in Clear Cell Renal Cell Carcinoma	166
7.3	Hypothesis-Free Prediction	166
7.4	Summary	167
<b>8.</b>	<b>Tissue Phenomics for Diagnostic Pathology</b>	<b>175</b>
	<i>Maria Athelougou and Ralf Huss</i>	
8.1	Introduction	175
8.2	Digital Pathology	176
8.3	Tissue Phenomics Applications for Computer-Aided Diagnosis in Pathology	178

8.3.1	Technical Prerequisites for Tissue Phenomics	180
8.4	Tissue Phenomics Applications for Decision Support Systems in Pathology	180
8.5	Summary of Pathology 4.0	182
<b>9.</b>	<b>Digital Pathology: Path into the Future</b>	<b>185</b>
	<i>Peter D. Caie and David J. Harrison</i>	
9.1	Introduction	186
9.2	Brief History of Digital Pathology	187
9.3	Digital Pathology in Current Practice	188
9.3.1	Research	188
9.3.2	Education	189
9.3.3	Clinical	190
9.4	Future of Digital Pathology	191
9.4.1	Mobile Scanning	191
9.4.2	Feature-Based Image Analysis	192
9.4.3	Machine Learning on Digital Images	195
9.4.4	Big Data and Personalized Pathology	196
<b>10.</b>	<b>Tissue Phenomics in Clinical Development and Clinical Decision Support</b>	<b>199</b>
	<i>Florian Leiß and Thomas Heydler</i>	
10.1	Cancer and Oncology Drug Development	199
10.2	Immunotherapy	200
10.3	Digitalization of Healthcare	202
10.4	Patient Profiling	203
10.5	Therapy Matching	205
10.6	Benefits	207
10.7	Conclusion	208
	<b>Glossary</b>	<b>209</b>
	<i>Index</i>	215

# Foreword

## Evolution of Tissue Phenomics and Why It Is Critical to the War on Cancer

### A view from a tumor immunologist and cancer immunotherapist

Those of us in biomedical research are witnessing an almost daily evolution of our science. Nowhere is this more obvious, or possessing greater impact, than in the field of cancer immunology and immunotherapy. Cancer, one of the great scourges on humanity, is having the veil of its secrets lifted. Digital imaging and objective assessment tools contribute substantial and solid evidence to document that immune cells are prognostic biomarkers of improved outcomes for patients with cancer. While anecdotal reports of associations between immune cell infiltrates and improved outcomes have been presented by pathologists for more than 100 years, the co-evolution of multiple science subspecialties has resulted in opportunities to better understand the disease and why it develops. Armed with this knowledge and evidence that checkpoint blockade therapies are capable of unleashing the immune system, increasing survival and possibly curing some patients with cancer, will lead to additional investment in this area of research, which will accelerate the pace at which we develop improved treatments for cancer. It is clear that digital imaging and assessment of complex relationships of cells within cancer, the very essence of tissue phenomics will play a central role in the development of the next generation of cancer immunotherapies. Ultimately, assessment of cancer tissue phenomics will be used to tailor immunotherapies to treat and eventually cure patients with cancer.

This is a very different time from when I began as an immunologist. Monoclonal antibodies, reagents capable of objectively assessing thousands of molecules, were not yet invented. To identify a subset of white blood cells, termed T cells, we used the binding of sheep erythrocytes to lymphocytes as the assay to characterize their numbers. The number of lymphocytes that formed *rosettes* was counted on a hemocytometer using a light microscope. This method was used to dose patients with immunosuppressive therapy to pre-

vent allograft rejection. In tissue sections and smears, we did not use immunohistochemistry (IHC); we used ocular annotation of a cell's morphological characteristics. My limited training in this area came at the hands of Dr. John W. Rebeck, a hematopathologist, who trained with Dr. Hal Downey, who trained with Professor Artur Pappenheim, at the University of Berlin. Professor Pappenheim, who developed the Pappenheim stains, educated his trainees into the subtleties of morphology, who then propagated the method to their trainees, and in this way the method spread.

Lymphocyte was my cell of interest; it was known to be a small round cell, with limited cytoplasm. Under Dr. Rebeck's tutelage, I denuded my skin with a scalpel, placed a drop of the diphtheria, pertussis, and tetanus vaccine on the area, covered it with a sterile glass coverslip, and attached it in place using a small piece of cardboard and adhesive tape. I would then change the coverslips every 3 h over a period of 48 h. Once the coverslips were removed and stained with Leishman's, I would sit at a multi-headed microscope and Dr. Rebeck would point out the monocytes or lymphocytes that were migrating into the site and onto the coverslip, identifying whether they were lymphocytes or monocytes that were morphing into large phagocytic cells. To support his description, Dr. Rebeck would describe the nuclear and cytoplasmic characteristics, as well as the other types of cells present in the area. That type of detailed evaluation has gone on for more than a century and remains the principal means to characterize disease.

As you read this book, it will become clear how the advances in image assessment technology allow subtle characteristics of cells to be evaluated in an objective and automated fashion. In addition to the morphological characteristics of cells, multiplex IHC provides simultaneous assessment of six or more markers on a single slide. Utilizing other technology, it is possible to stain a slide and then image and strip the slide of the reagents so that the cycle can be repeated as many as 60 times, allowing assessment of as many markers on a single slide. Coupled with the advent of tissue phenomics, all of this information can be evaluated in the context of whether it is inside the tumor, at the invasive margin, or in the stroma.

The molecular evaluation of disease is also advancing. Summaries of gene expression profiling data for tumor samples from hundreds of patients are available in The Cancer Genome

Atlas (TCGA) and other databases. These databases are being interrogated to evaluate how many cancers had high or low levels of genes associated with immune cells, as well as for expression of cancer-destroying molecules or of mechanisms cancer can use to evade immune-mediated destruction. While these interrogations of the data are providing important and powerful insights about the immune system's response to cancer, much of this information lacks the context of where these elements are expressed and the identity of which cells are expressing specific genes. Only by understanding the context of this information, specifically which cell is expressing what and which cells are nearby, can it be effectively used to guide a new generation of cancer immunotherapy trials.

Head and neck cancer will serve as an example to further clarify this point. For several years, it has been known that increased numbers of CD8(+) *cancer killer T cells* at the tumor were associated with improved survival. For example, in one study patients whose tumors had above the median number of CD8 T cells had around a 50% 5-year survival, while those whose tumors had CD8 T cell numbers below the median, had a 35% 5-year survival. This pattern was true for assessment of CD8(+) T cell numbers by IHC or gene expression profiling. As an immunologist who recognizes the important role that CD8 T cells can play in preclinical animal models, this made sense. However, it was also reported that an increased number of FOXP3(+) *suppressor T cells* were also associated with improved survival. Since these are the cells that can turn off the cancer killer cells, these results made no sense. As we began to apply the multiplex IHC method to visualize six markers on a single 4-micron section, it became clear that in some patients, immune cells associated with the tumor were organized in a specific pattern. In one case, the tumor, which uniformly expressed high levels of the immune checkpoint PD-L1, had excluded essentially all immune cells from inside the tumor. However, at the tumor-stroma barrier was a band of suppressor T cells, and outside of that band were the CD8(+) cancer killer T cells. Since several of the suppressor cells' immune inhibitory functions required cell contact, we reasoned that in order to be effective, the suppressor cells would need to be relatively close to the CD8(+) cancer killer cell that they were trying to inhibit. When we evaluated tumors with a high number of suppressor cells near the CD8(+) cancer killer cells, we found that these patients did significantly worse than patients with low numbers, not better. As

we evaluated another inhibitory molecule, PD-L1, which mediates inhibition by contact, we found the same pattern. While additional validation needs to be done, this illustrates the power of evaluating cell–cell relationships. It will be these types of assessments with panels of 20–25 markers that will provide critical insights into why tumors escape immune elimination and what hurdles will need to be addressed to improve patient outcomes. It is not going to be easy and after spending 5 years in the midst of digital imaging technologies, I realize there are substantial challenges ahead. There is no looking back! Since the immune system is the critical element that can cure patients of cancer, it is essential to assess the cancer that escapes and ultimately kills the patient. While multiple approaches must be applied, the *tissue is the issue* and tissue phenomics is the approach that will play the critical role in unraveling the amazing complexity of the cancer–immune system interphase and drive the development of treatments to cure patients with cancer.

**Bernard A. Fox**

Providence Cancer Center

Portland, OR, USA

Autumn 2017

# Acknowledgments

We express our gratitude to all co-authors and contributors, who worked with great commitment and passion on the successful completion of the different chapters and the finishing of the entire book.

We also thank everybody in Definiens AG for insightful and astute discussions to provide a comprehensive and coherent overview on *tissue phenomics* and related subjects, such as cognition networks, assay development, digital pathology, image analysis, big data analytics, and machine learning. We very much appreciate the generous support of Definiens AG in providing us leeway and resources to create this book.

A very special thank goes to Ralf Schönmeier for his enthusiastic organizational support, manuscript editing, and putting everything together in the desired format.

We thank the members of our advisory board for their devoted leadership and inspirational guidance and our investors and shareholders for their solid confidence and trust in such a disruptive technology and our creative people.

We are grateful to our longtime academic collaborators, business partners, and commercial customers for their outstanding insights and challenging thoughts. This holds true in particular for the attendees of our annual *Tissue Phenomics Symposium* and their enthusiasm for novel ideas and creative solutions.

And finally, we thank Pan Stanford Publishing for inviting us to write this book, contribute to the editorial work, and for subsequently publishing it.

**Gerd Binnig**  
**Ralf Huss**  
**Günter Schmidt**

