

Preface

In his superb exposition, *The Emperor of All Maladies: A Biography of Cancer*, Mukherjee attributes the earliest documentation of cancer to the brilliant Egyptian, Imhotep, who some 4,500 years ago clearly described a case of breast cancer (Mukherjee 2010). Roughly two millennia later (ca. 400 BC), the Greek physician Hippocrates named the disease *karkinos* (the Greek word for crab), which has now come down to us as cancer. Some five to six centuries later while practicing in Rome (ca. 130–200 AD), the Greek physician, Claudius Galen, who was influenced by the four humors constituting the human body as proposed by the Hippocratic school, i.e., blood, phlegm, yellow bile, and black bile, attributed cancer to an excess of black bile. It took centuries before Vesalius (sixteenth century) and Baillie (eighteenth century) put the black bile hypothesis to rest, thus indirectly encouraging surgeons to begin resection of solid tumors. (Surgical procedures had been done earlier by some fearless surgeons, but few patients survived the ordeal and infection that likely followed.) The later introduction of anesthesia and antibiotics in the nineteenth to twentieth centuries, as well as more sterile operating environments, thrust surgery (and later radiation therapy) as a major treatment of this disease, an approach that is still used whenever possible. In the middle of the twentieth century and continuing today, chemotherapy and hormonal therapy emerged as a complement to, and sometimes instead of, surgery and radiation therapy to treat cancer.

A number of theories have been proposed regarding those factors that may drive and facilitate a cancer to initiate, develop, and metastasize, and these have guided cancer studies in the past few decades. An insightful speculation was made by Otto Warburg following his seminal work in the 1920s: “Cancer ... has countless secondary causes. But ... there is only one prime cause, [which] is the replacement of respiration of oxygen in normal body cells by a fermentation of sugar” (Warburg 1969).

The first discovery of oncogenes and tumor suppressor genes about 40 years ago marked another major milestone in our understanding of cancer development, which has profoundly influenced research in this area during the past three decades. It has become a widely held belief that cancer is ultimately a disease caused by genomic mutations. Aided by the rapidly increasing pool of a variety of *omic* data such as genomic, transcriptomic, epigenomic, metabolomic, glycomic, lipidomic,

and pharmacogenomic data collected on both cell lines and cancer tissues, spectacular progress has been made in the past two decades in our understanding of cancer, particularly in terms of how the microenvironment and the immune system contribute to the whole process of neoplasm formation and survival.

In spite of the considerable progress made, however, a number of salient questions remain to be answered. The authors posit that a considerable amount of information needed to address and answer many of these questions already exists in the available *omic* databases, and much of these data are substantially undermined and underutilized. Among the many possible reasons, a key one, we believe, is that computational cancer biologists, as a community, have yet to sufficiently develop their independent thinking about the overall biology of cancer. The thinking should be quite different from the reductionist approaches that have been widely used in experimental studies of cancer in the past century and should enable them to address fundamental questions about cancer in a holistic manner as an evolving system. Many fundamental issues concerning cancer are intrinsically holistic by nature. Thus, when examining cancer as an evolutionary problem, its microenvironment, including the extracellular matrix and the immune and other stromal cells, must be considered as an integral part of the system. This strongly suggests that cell culture-based or animal model-based cancer studies must be complemented by cancer tissue-based studies in order to gain a full understanding of cancer. The *omic* data collected on cancer tissue samples, covering different developmental stages, is likely to contain the information on the interplay between cancer cells and their environment, and particularly how such interactions may drive the evolution in specific directions. Hence, we posit that mining such *omic* data for information discovery will, in the future, represent an essential component of cancer research, complementary to the current more reductionist-oriented approaches.

The goals of this book are to provide an overview of cancer biology from an informatics perspective and to demonstrate how *omic* data can be mined to generate new insights and a more comprehensive understanding that is needed to address a wide range of fundamental cancer biology questions. Throughout this book, the authors have attempted to establish the following key points: (1) cancer is a process of cell survival in an increasingly more stressful and difficult microenvironment, which co-evolves with the diseased cells; (2) cell proliferation is a cancer's way to reduce the stresses imposed on them for survival; (3) the challenges that the evolving cells must overcome are not only at the cell level, but more importantly at the tissue level, hence making cancer dominantly a tissue rather than a cell-only problem; (4) the survival pathway for each cancer is not created 'on the fly' through its selection of molecular malfunctions or genetic mutations, instead it is largely determined by substantial cellular programs encoded in the human genome, which originally evolved for other purposes; (5) subpopulations of cancer cells have managed to create the conditions needed to trigger such cellular program-guided survival pathways; (6) as the stresses become increasingly more challenging, cancer cells utilize increasingly less reversible stress-responses for their survival, thus making the disease progressively more malignant; (7) genomic mutations in sporadic cancers probably serve mainly as permanent replacements for ongoing functions to

provide efficiency and sustainability for survival; in contrast, mutations in hereditary cancers dominantly play driver roles of cancer initiation, but in a sense different from driver mutations as defined in the current literature; (8) there is a fundamental difference between cell proliferation in primary *versus* metastatic cancers as the former is essential in overcoming the encountered stress(es) while the latter is simply a side product of a stress-response process, suggesting that their treatment regimens should be different; and (9) cancer survives and proliferates by continually evolving with natural selection having a major part in deciding which cells remain and which must perish.

For each chapter, the authors present the main topic by placing cancer in an evolutionary context, for example by raising and addressing questions such as: *What pressures are the evolving neoplastic cells currently under*, and *How have the cells responded to adapt to the pressures?* In addition, the authors also demonstrate through examples how to derive the desired information from the available *omic* data by asking questions and then addressing them using a hypothesis-driven data mining approach. An example could be as follows: *What is the difference between the main driving forces of primary versus metastatic cancer?* This can be addressed by identifying genes that are up-regulated consistently across all metastatic cancers *versus* their matching primary cancer tissues, and then delineating the particular pathways that are enriched by these genes.

This 14-chapter book consists of the following clusters of chapters. Chapters 1 and 2 introduce the basic biology and biochemistry of cancer and the available cancer *omic* data, as well as the type of information derivable from such data. Chapter 3 serves as an introduction to the use of *omic* data to address cancer-related problems, written for someone with only a limited knowledge of cancer; and Chap. 12 serves a similar purpose but for someone who has a general understanding about cancer at the molecular and cellular levels, e.g., having read a substantial portion of this book. Chapter 4 is a transition chapter, serving as an introduction to both information that can be derived from cancer genomes and elucidation of cancer mechanisms using such information. Chapters 5 through 9 represent the core of the book: elucidation of novel information and how to gain a new and better understanding about the fundamental biology of primary cancer, in which cancer is treated as an evolving system driven by specific pressures and assisted by certain facilitators at different developmental stages. A common theme is used when tackling a series of cancer-related key issues across these five chapters: *What stresses do the cancer cells need to overcome at a specific stage*, and *how do such cells utilize encoded stress-response systems to ensure their survival?* Chapters 10 and 11 extend this discussion to metastatic cancer, which, somewhat surprisingly, represents a different type of disease from primary cancers with fundamentally different drivers. Chapter 13 provides some general information to those new to the field about how to conduct meaningful data mining-based cancer research. Chapter 14 presents our perspectives about cancer research using a more holistic approach than is generally done.

The authors hope that this book will help in bridging the gap between experimental cancer biologists and computational biologists in their joint efforts to uncover the

enormous wealth of information hidden in the cancer *omic* data. Success in this endeavor will lead to a better understanding of cancer, as well as assist computational biologists to develop independent thinking when tackling these complex problems. This approach will probably be less detail-oriented but more holistic and will likely span the entire range of cancer evolution, thus making it different from but complementary to those of their experimental peers. It is the authors' contention that more qualitative and quantitative utilization of the *omic* data will improve our overall understanding of cancer biology, hence leading to improved capabilities in early detection, development of more effective cancer treatments, and improvement in the quality of the patient lives.

The authors welcome any feedback from the reader regarding errors that need correcting and areas where the book could be improved. Such information will be highly valuable, particularly if there is a decision to write a future edition of the book.

Athens, GA, USA
Lincoln, NE, USA
Chapel Hill, NC, USA

Ying Xu
Juan Cui
David Puett

References

- Mukherjee, S. (2010). The emperor of all maladies: a biography of cancer, Scribner.
Warburg, O. H. (1969). The prime cause and prevention of cancer, K. Triltsch.

Cancer Bioinformatics

Xu, Y.; Cui, J.; Puett, D.

2014, XXVI, 368 p. 68 illus., Hardcover

ISBN: 978-1-4939-1380-0