Cancer Informatics in the Post Genomic Era

Toward Information-Based Medicine

Cancer Treatment and Research

Steven T. Rosen, M.D., *Series Editor*

Arnold, A.J. (ed.): *Endocrine Neoplasms*. 1997. ISBN 0-7923-4354-9.
Pollock, R.E. (ed.): *Surgical Oncology*. 1997. ISBN 0-7923-9900-5.
Verweij, J., Pinedo, H.M., Suit, H.D. (eds): *Soft Tissue Sarcomas: Present Achievements and Future Prospects.* 1997. ISBN 0-7923-9913-7.
Walterhouse, D.O., Cohn, S. L. (eds): *Diagnostic and Therapeutic Advances in Pediatric Oncology*. 1997. ISBN 0-7923-9978-1.
Mittal, B.B., Purdy, J.A., Ang, K.K. (eds): *Radiation Therapy*. 1998. ISBN 0-7923-9981-1.
Foon, K.A., Muss, H.B. (eds): *Biological and Hormonal Therapies of Cancer*. 1998. ISBN 0-7923-9997-8.
Ozols, R.F. (ed.): *Gynecologic Oncology*. 1998. ISBN 0-7923-8070-3.
Noskin, G. A. (ed.): *Management of Infectious Complications in Cancer Patients*. 1998. ISBN 0-7923-8150-5.
Bennett, C. L. (ed.): *Cancer Policy*. 1998. ISBN 0-7923-8203-X.
Benson, A. B. (ed.): *Gastrointestinal Oncology*. 1998. ISBN 0-7923-8205-6.
Tallman, M.S., Gordon, L.I. (eds): *Diagnostic and Therapeutic Advances in Hematologic Malignancies*. 1998. ISBN 0-7923-8206-4.
von Gunten, C.F. (ed.): *Palliative Care and Rehabilitation of Cancer Patients*. 1999. ISBN 0-7923-8525-X
Burt, R.K., Brush, M.M. (eds): *Advances in Allogeneic Hematopoietic Stem Cell Transplantation*. 1999. ISBN 0-7923-7714-1.
Angelos, P. (ed.): *Ethical Issues in Cancer Patient Care* 2000. ISBN 0-7923-7726-5.
Gradishar, W.J., Wood, W.C. (eds): *Advances in Breast Cancer Management*. 2000. ISBN 0-7923-7890-3.
Sparano, J. A. (ed.): *HIV & HTLV-I Associated Malignancies*. 2001. ISBN 0-7923-7220-4.
Ettinger, D. S. (ed.): *Thoracic Oncology*. 2001. ISBN 0-7923-7248-4.
Bergan, R. C. (ed.): *Cancer Chemoprev*ention. 2001. ISBN 0-7923-7259-X.
Raza, A., Mundle, S.D. (eds): *Myelodysplastic Syndromes & Secondary Acute Myelogenous Leukemia* 2001. ISBN: 0-7923-7396.
Talamonti, M. S. (ed.): *Liver Directed Therapy for Primary and Metastatic Liver Tumors*. 2001. ISBN 0-7923-7523-8.
Stack, M.S., Fishman, D.A. (eds): *Ovarian Cancer*. 2001. ISBN 0-7923-7530-0.
Bashey, A., Ball, E.D. (eds): *Non-Myeloablative Allogeneic Transplantation*. 2002. ISBN 0-7923-7646-3.
Leong, S. P.L. (ed.): *Atlas of Selective Sentinel Lymphadenectomy for Melanoma, Breast Cancer and Colon Cancer.* 2002. ISBN 1-4020-7013-6.
Andersson , B., Murray D. (eds): *Clinically Relevant Resistance in Cancer Chemotherapy*. 2002. ISBN 1-4020-7200-7.
Beam, C. (ed.): *Biostatistical Applications in Cancer Research*. 2002. ISBN 1-4020-7226-0.
Brockstein, B., Masters, G. (eds): *Head and Neck Cancer*. 2003. ISBN 1-4020-7336-4.
Frank, D.A. (ed.): *Signal Transduction in Cancer*. 2003. ISBN 1-4020-7340-2.
Figlin, R. A. (ed.): *Kidney Cancer*. 2003. ISBN 1-4020-7457-3.
Kirsch, M.; Black, P. McL. (ed.): *Angiogenesis in Brain Tumors*. 2003. ISBN 1-4020-7704-1.
Keller, E.T., Chung, L.W.K. (eds): *The Biology of Skeletal Metastases.* 2004. ISBN 1-4020-7749-1.
Kumar, R. (ed.): *Molecular Targeting and Signal Transduction*. 2004. ISBN 1-4020-7822-6.
Verweij, J., Pinedo, H.M. (eds): *Targeting Treatment of Soft Tissue Sarcomas*. 2004. ISBN 1-4020-7808-0.
Finn, W.G., Peterson, L.C. (eds.): *Hematopathology in Oncology*. 2004. ISBN 1-4020-7919-2.
Farid, N. (ed.): *Molecular Basis of Thyroid Cancer*. 2004. ISBN 1-4020-8106-5.
Khleif, S. (ed.): *Tumor Immunology and Cancer* Vaccines. 2004. ISBN 1-4020-8119-7.
Balducci, L., Extermann, M. (eds): *Biological Basis of Geriatric Oncology*. 2004. ISBN
Abrey, L.E., Chamberlain, M.C., Engelhard, H.H. (eds): *Leptomeningeal Metastases*. 2005. ISBN 0-387-24198-1
Platanias, L.C. (ed.): *Cytokines and Cancer*. 2005. ISBN 0-387-24360-7.
Leong, S.P.L., Kitagawa, Y., Kitajima, M. (eds): *Selective Sentinel Lymphadenectomy for Human Solid Cancer.* 2005. ISBN 0-387-23603-1.
Small, Jr. W., Woloschak, G. (eds): *Radiation Toxicity: A Practical Guide*. 2005. ISBN 1-4020-8053-0.
Haefner, B., Dalgleish, A. (eds): *The Link Between Inflammation and Cancer*. 2006. ISBN 0-387-26282-2.
Leonard, J.P., Coleman, M. (eds): *Hodgkin's and Non-Hodgkin's Lymphoma.* 2006. ISBN 0-387-29345.
Leong, S.P.L. (ed): *Cancer Clinical Trials: Proactive Strategies*. 2006. ISBN 0-387-33224-3.
Meyers, C. (ed): *Aids-Associated Viral Oncogenesis.* 2007. ISBN 978-0-387-46804-4.
Ceelen, W.P. (ed): *Peritoneal Carcinomatosis: A Multidisciplinary Approach.* 2007. ISBN 978-0-387-48991-9.
Leong, S.P.L. (ed): *Cancer Metastasis and the Lymphovascular System: Basis for rational therapy*. 2007. ISBN 978-0-387-69218-0.
Raizer, J., Abrey, L.E. (eds): *Brain Metastases.* 2007. ISBN 978-0-387-69221-0.
Jurisica, I., Wigle, D.A., Wong, B. (eds): *Cancer Informatics in the Post Genomic Era.* 2007. ISBN 978-0-387-69320-0

Cancer Informatics in the Post Genomic Era

Toward Information-Based Medicine

edited by

Igor Jurisica, PhD
Ontario Cancer Institute, PMH/UHN
Toronto Medical Discovery Tower
Toronto, Ontario

Dennis A. Wigle, MD, PhD
Division of Thoracic Surgery, Mayo Clinic
Mayo Clinic Cancer Center
Rochester, Minnesota, USA

Bill Wong, BSc, MBA
Program Director
Information Management
IBM Toronto Laboratory
Markham, Ontario

Springer

Igor Jurisica, PhD
Ontario Cancer Institute, PMH/UHN
Toronto Medical Discovery Tower
Division of Signaling Biology
Life Sciences Discovery Centre
Room 9-305, 101 College Street
Toronto, Ontario M5G 1L7 CANADA

Dennis A. Wigle, MD
Division of Thoracic Surgery
Mayo Clinic Cancer Center
200 First St. SW
Rochester, Minnesota 55905 USA

Bill Wong
Database Competitive Technologies
IBM Toronto Laboratory
8200 Warden Avenue
Markham, ON L3R 9Z7 CANADA

Series Editor:
Steven T. Rosen
Robert H. Lurie Comprehensive Cancer Center
Northwestern University
Chicago, IL
USA

Cancer Informatics in the Post Genomic Era: Toward Information-Based Medicine

Library of Congress Control Number: 2006939420

ISBN-10: 0-387-69320-3 e-ISBN-10: 0-387-69321-1
ISBN-13: 978-0-387-69320-0 e-ISBN-13: 978-0-387-69321-7

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

Table of Contents

List of Figures

Foreword

The healthcare and pharmaceutical industries have been "buzzing" with the promise of personalized healthcare since the inception of the human genome project. Information technology will accelerate the delivery of advances in medical science and technology to the public. How will the convergence of information technology and life sciences impact the future?

During the past decade, life sciences and information technology began to converge, resulting in significant and life-impacting research – the result with perhaps the highest impact to date being the sequencing of the human genome and its influence on how clinical researchers now investigate methods and molecules that could improve the human condition. Knowledge gained through human genome sequencing is driving recent achievements in genomics, proteomics, molecular biology and bioinformatics. As the decade progresses, next generation medical science technology and capabilities, enabled by increasingly "smarter" information technology, will change the pace of discovery, development and delivery of new treatments even more dramatically. For example, biopharmaceutical research will continue to shift from a small, molecule-centered approach to one of stronger biomedical emphasis. This shift will focus on moving from the molecular actions of small molecule compounds toward delivering biological-based diagnostics and therapeutics. Healthcare will become increasingly personalized as these biological-based diagnostics and treatments become standard practice.

The application of information technology advances to those discoveries in science and medicine is giving rise to a new discipline, information-based medicine, which provides new knowledge by integrating and analyzing data from patients' clinical information,

medical images, the environment, genetic profiles, as well as molecular and genomic research efforts. Information-based medicine is the marriage of information technology with the practice of medicine and pharmaceutical research for improved disease diagnosis, therapeutics and healthcare delivery. Information-based medicine is the use of information technology to achieve personalized medicine.

From developing health information networks for nations around the world to being a founding member of the Worldwide Biobank Summit, IBM is driving innovation in the healthcare and life sciences industries. IBM welcomes books like this that advances the industry's move toward information-based medicine and targeted treatment solutions.

Michael Svinte
Vice President of Information Based Medicine
IBM Healthcare and Life Sciences

Preface

Less than 50% of diagnosed cancers are cured using current treatment modalities. Many common cancers can already be fractionated into such therapeutic subsets with unique prognostic outcomes based on characteristic molecular phenotypes. It is widely expected that treatment approaches of complex cancer will soon be revolutionized by combining molecular profiling and computational analysis, which will result in the introduction of novel therapeutics and treatment decision algorithms that target the underlying molecular mechanisms of cancer.

The sequencing of the human genome was the first step in understanding the ways in which we are wired. However, this genetic blueprint provides only a "parts list", and neither information about how the human organism is actually working, nor insight into function or interactions among the ~30 thousand constitutive parts that comprise our genome. Considering that the 30 years of worldwide molecular biology efforts have only annotated about 10% of this gene set, and we know even less about proteins, it is comforting to know that high-throughput data generation and analysis is now widely available.

By arraying tens of thousands of genes and analyzing abundance of and interaction among proteins, it is now possible to measure the relative activity of genes and proteins in normal and diseased tissue. The technology and datasets of such profiling-based analyses will be described along with the mathematical challenges that face the mining of the resulting datasets. We describe the issues related to using this information in the clinical setting, and the future steps that will lead to drug design and development to cure complex diseases such as cancer.

Contributing Authors

Igor Jurisica, PhD

Dr. Jurisica is a Canada Research Chair in Integrative Computational Biology, a Scientist at the Ontario Cancer Institute, University Health Network since 2000, Associate Professor in the Departments of Computer Science and Medical Biophysics, University of Toronto, Adjunct Professor at School of Computing Science, Queen's University, and a Visiting Scientist at the IBM Centre for Advanced Studies. He earned his Dipl. Ing. degree in Computer Science and Engineering from the Slovak Technical University in 1991, M.Sc. and Ph.D. in Computer Science from the University of Toronto in 1993 and 1998 respectively.

Dr. Jurisica's research focuses on computational biology, and representation, analysis and visualization of high dimensional data generated by high-throughput biology experiments. Of particular interest is the use of comparative analysis for the mining of integrated datasets such as protein—protein interaction, gene expression profiling, and high-throughput screens for protein crystallization.

Scientist
Ontario Cancer Institute, PMH/UHN
Toronto Medical Discovery Tower
Division of Signaling Biology
Life Sciences Discovery Centre
Room 9-305
101 College Street
Toronto, Ontario M5G 1L7

Tel./Fax: 416-581-7437
Email: juris@ai.utoronto.ca
URL: http://www.cs.utoronto.ca/~juris

Associate Professor
Departments of Computer Science and Medical Biophysics,
University of Toronto

Dennis A. Wigle, MD, PhD

Since August 2006, Dennis Wigle has been a clinician-scientist at the Mayo Clinic Cancer Center in Rochester Minnesota. He is a practicing thoracic surgeon with an interest in thoracic oncology. His laboratory investigates the genetic basis and molecular sequence of events underlying thoracic malignancies. He holds an MD from the University of Toronto and a PhD from the Department of Anatomy and Cell Biology at Queen's University in Kingston, Canada. His interests include the application of novel computational methods to the analysis of high-throughput data in cancer biology.

Dennis A. Wigle
Division of Thoracic Surgery, Mayo Clinic
Mayo Clinic Cancer Center
200 First St. SW Rochester, Minnesota USA 55905

Tel: (507) 284-4099 (clinical office)
Tel: (507) 284-8462 (secretary)
Tel: (507) 538-0558 (lab office)
Fax: (507) 284-0058
Email: wigle.dennis@mayo.edu

Bill Wong

Bill Wong has an extensive background is software deployment technologies and has been working with a variety of database technologies. Some of his previous roles included being the Information Management product manager for Life Sciences, Linux, and Grid solutions. His current role is Program Director for Advanced Database Technologies at IBM. He works out of the

Toronto Lab and can often be found speaking at conferences on information management future trends and directions.

Program Director
Database Competitive Technologies
IBM Toronto Laboratory
8200 Warden Avenue
Markham, ON L3R 9Z7

Tel.: 905 413-2779, Fax: 905 413- 4928 T/L: (969)
Email: billw@ca.ibm.com

Thomas Kislinger, PhD

Dr. Kislinger is a Canada Research Chair in Proteomics in Cancer Biology, a Scientist at the Ontario Cancer Institute, University Health Network since 2006 and an Assistant Professor in the Department of Medical Biophysics at the University of Toronto. He earned his M.Sc. equivalent in Analytical Chemistry from the Ludwig-Maximilians University in Munich, Germany (1998) and his Ph.D. in Analytical Chemistry from the Friedrich-Alexander University in Erlangen, Germany and the Columbia University, New York (2001). He carried out his post-doctoral research at the Banting & Best Department of Medical Research in Toronto where he developed an expertise in large-scale expression proteomics of mammalian model systems.

Dr. Kislinger's research interests are focused on the development and application of shot-gun proteomics to diverse question in cancer, vascular and cardiovascular biology.

Scientist
Ontario Cancer Institute
MaRS Centre
Toronto Medical Discovery Tower
9th floor Room 9-807
101 College Street
Toronto, Ontario
Canada M5G 1L7

Telephone: 416-581-7627
Fax: 416-581-7629
e-mail: thomas.kislinger@utoronto.ca
URL: http://medbio.utoronto.ca/faculty/kislinger.html

Assistant Professor
Department of Medical Biophysics, University of Toronto

Ming-Sound Tsao, MD

Dr. Tsao is the M. Qasim Choksi Chair in Lung Cancer Translational Research and Professor of Laboratory Medicine and Pathobiology at University of Toronto. He is a Senior Scientist and Surgical Pathologist with special interest in neoplastic diseases of the aerodigestive tract. His focus on lung cancer research is on the identification and validation of molecular prognostic markers for early stage lung cancer patients, especially using genome-wide expression and genomic microarray platforms. He is also interested in predictive markers for benefits from adjuvant chemotherapy and targeted therapy in lung cancer. He has published more than 160 peer-reviewed manuscripts, with the most recent one published in the New England Journal of Medicine on molecular and clinical predictors of outcome in lung cancer patients treated by erlotinib.

Senior Scientist
Ontario Cancer Institute
Princess Margaret Hospital
610 University Avenue
Toronto Ontario M5G 2M9 Canada

Tel.: 416-340-4737
Email: ming.tsao@uhn.on.ca

Professor
Department of Laboratory Medicine and Pathobiology,
Univ. of Toronto

Chunlao Tang, PhD

Dr. Tang earned his PhD from the Molecular and Computational Biology Program at the University of Southern California. His primary research interest lies in studying the genetic basis underlying natural phenotypic variation. He seeks to elucidate the genetic variation associated with susceptibility to common diseases in humans using genomic approaches.

Email: chunlaot@usc.edu

Acknowledgments

Igor Jurisica gratefully acknowledges his lab members who contributed to the re-search results and stimulating research environment, especially Kevin Brown, Baiju Devani, Michael McGuffin, David Otasek, Mahima Agochiya, Dan Strumpf, Frederic Breard, Richard Lu and other students and programmers. Spe-cial thanks to numerous collaborators in lung, ovarian, prostate cancer sites.

Bill Wong would like to make the acknowledgements to the following people from IBM for their contributions: Mike Svinte and the Healthcare and life Sci-ences marketing staff, the IBM Business Consulting Services authors of the Per-sonalized Healthcare 2010 - Are you ready for information-based medicine paper, Barbara Eckman and Douglas Del Prete for the SQL queries, Richard Hale for his contributions regarding online analytical processing and data mining.

Dennis Wigle would like to thank lab members and members of the research community at Mayo Clinic who continue to inspire innovative approaches to problems in thoracic oncology.