Valeria Zazzu · Maria Brigida Ferraro
Mario R. Guarracino *Editors*

Mathematical Models in Biology

Bringing Mathematics to Life

Springer

Mathematical Models in Biology

Valeria Zazzu • Maria Brigida Ferraro
Mario R. Guarracino
Editors

Mathematical Models in Biology

Bringing Mathematics to Life

Springer

Editors
Valeria Zazzu
Institute of Genetics and Biophysics
(IGB) "ABT"
National Research Council of Italy (CNR)
Naples, Italy

Mario R. Guarracino
High Performance Computing
and Networking Institute (ICAR)
National Research Council of Italy (CNR)
Naples, Italy

Maria Brigida Ferraro
Department of Statistical Sciences
Sapienza University of Rome
Rome, Italy

ISBN 978-3-319-23496-0 ISBN 978-3-319-23497-7 (eBook)
DOI 10.1007/978-3-319-23497-7

Library of Congress Control Number: 2015956528

Mathematics Subject Classification (2010): 92c99,92Bxx, 97M10, 90B10, 90B15, 92C42, 92B20

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

Life scientists are not always fully aware of the powerful role that mathematical models have in both answering biological research questions and in making predictions. Scientists have a clear view of the problem; they know the questions; they have identified ways to answer; and they produce the data to be analysed. Novel high throughput technologies are utilized that give rise to an unprecedented quantity of data. However, the data is 'noisy', and the answer to each question can be well hidden under terabytes of incomprehensible text files.

It is here that the mathematicians can help: they know 'how' to do things; they love the huge, ugly text files; they foresee hundreds of statistics that could be calculated; they want to try all of them because there is always uncertainty. Mathematicians see paths, trends, connections, and correlations. Ultimately the need to identify the beautiful biological mechanisms that are hidden, must come to light. Indeed, mathematicians too, get stuck, lost among protein sticks, bubbles, helices, and sheets.

During the 'Bringing Maths to Life' workshop, held in Naples, Italy, October 27–29, 2014, biologists and mathematicians joined forces to address key areas in biology that face demanding mathematical challenges. A list of invited speakers and participants came from leading European universities and the international scientific community; especially computational biologists, mathematicians, and researchers in the life sciences. Interdisciplinary discussions surrounded existing cases in an effort to identify gaps or to share existing solutions. Finding the best mathematical resolution to interpret data from a biological perspective, or—inversely—understanding the biological issue and its real-life constraints from a mathematical viewpoint, required both communities to closely engage. The present volume gathers a number of chapters selected from the most interesting contributions to the workshop.

The workshop had featured three main sessions. '**Zoom inside the cell: microscopy images processing**' had been the topic of the first session. Biological visualization provides the means through which to place genomic and proteomic information in a cellular or tissue context. While existing software enables particular assays for distinct cell types, high throughput image analysis has, to this point, been impractical unless an image analysis expert develops a customized solution, or

unless commercial packages are used with their built-in algorithms for limited sets of cellular features and cell types. There exists a clear need for powerful, flexible tools for high throughput cell image analysis. Computer vision researchers have contributed new algorithms to the project so that their theoretical work can be applied to practical biological problems.

The session on '**Genetic variability and differential expression: sequence data analysis**' had addressed the recent revolution in DNA sequencing technology brought by the sequencing of an increasing number of genomes. Changes in data quantity and format (large numbers of short reads or pairs of short reads *versus* relatively long reads produced by traditional Sanger sequencing) imply changes of sequence data management, storage, and visualization, and provide a challenge for bioinformatics.

'**Deciphering complex relationships: networks and interactions**' had dealt with biological systems composed of thousands of different types of components and the problems related to the huge networks that comprise numerous non-linearly interacting dimensions, from which, in turn, biological functions emerge. The networks are far too complex to be understood by the unassisted human mind and therefore to analyze these complex biological systems and to obtain relevant answers, biology requires quantitative models that draw from modern computer science and mathematics.

Additionally, there had been three invited sessions. The first one was on '**Molecular Dynamics and Modelling of Protein Structure and Function via High Performance Computing Simulations**' (organized by Alessandro Grottesi from CINECA, Italy). Molecular dynamics simulations are computational tools aimed at studying protein structure and dynamics as well as protein-protein interactions at the atomic level. The high performance computing of current computer architectures, as well as the developing of valid force fields for the mathematical modelling of biochemical interactions, have provided new tools to help biologists studying and testing hypotheses to understand biochemical phenomena in a new perspective. This session has highlighted the advantages and limitations of this powerful computational technique.

In the second invited session, '**Statistical challenges in omics research within Life Sciences**' (organized by J.J. Houwing-Duistermaat from Leiden University Medical Center, The Netherlands and Luciano Milanesi from Institute of Biomedical Technologies, CNR, Italy), several statistical issues in omics datasets were addressed, from preprocessing up to building statistical models for joint interpretation of the datasets. These datasets contain information about different aspects of the same biological processes. Therefore in many studies, multiple omics datasets are nowadays available and integrated analyses of these omics datasets is the ultimate goal to understand biological mechanisms underlying traits. However integration of these datasets is not straightforward since they vary in measurement error distributions, scale, sparseness and size. In this session challenges were addressed in single omics datasets analysis as well as combined analysis of multiple omics datasets.

The third invited session had been dedicated to '**Artificial neurons and realistic simulation of neuronal functions**' (organized by Angela Tino from Institute of Cybernetics, CNR, Italy). Modern neuroscience research has generated vast volumes of experimental data, and large scale initiatives launched in recent years will gather much more. Nonetheless, much of the knowledge needed to build multilevel atlases and unifying models of the brain is still missing. Brains are a large network composed of many neurons with their synaptic connections, each expressing different proteins on the cell membrane and each with its own complex internal structure. Despite huge advances, there is no technology that allows us to characterize more than a tiny part of this complexity. The session had shed light on novel solutions from neural-inspired artificial models and software, realistic neuronal function simulation, and functional and molecular neurobiology and had aimed to gather scientists from diverse disciplines to foster integrated approaches to unravel complex brain functions.

Naples, Italy — Valeria Zazzu
Rome, Italy — Maria Brigida Ferraro
Naples, Italy — Mario R. Guarracino

Acknowledgments

The workshop has been organized by: Alessandra Rogato (Institute of Biosciences and Bioresources); Valeria Zazzu and Enza Colonna (Institute of Genetics and Biophysics); Mario Guarracino (High Performance Computing and Networking Institute and Institute for Higher Mathematics 'F. Saveri') from the Italian National Research Council (CNR), Italy; Maria Brigida Ferraro from Sapienza University of Rome, Italy; Martijn Moné from the VU University Amsterdam and ISBE—Infrastructure for Systems Biology Europe, The Netherlands. Gerardo Toraldo from the Department of Mathematics and Applications 'Renato Caccioppoli', University of Naples Federico II, contributed to the organization.

The initiative has been supported by the Italian National Research Council (CNR), the Institute for High Mathematics 'F. Saveri' (INDAM), the High Performance Computing and Networking Institute (ICAR), the University of Naples Federico II, Mimomics, Interomics, the Department of Bio-Agriculture Sciences (CNR), the Department of Biomedical Sciences (CNR), LABGTP and Tecnologica.

Contents

Image Segmentation, Processing and Analysis in Microscopy and Life Science

Carolina Wählby

Abstract Microscopes have been used for more than 400 years to understand biological and biomedical processes by visual observation. Science is the art of observing, but science also requires measuring, or quantifying, what is observed. Research based on microscopy image data therefore calls for methods for quantitative, unbiased, and reproducible extraction of meaningful measurements describing what is observed. Digital image processing and analysis is based on mathematical models of the information contained in image data, and allows for automated extraction of quantitative measurements. Automated methods are reproducible and, if applied consistently and accurately across experiments with positive as well as negative controls, also unbiased. Digital image processing is further motivated by the development of scanning microscopes and digital cameras that can capture image data in multiple spatial-, time-, and spectral-dimensions, making visual assessment cumbersome or even impossible due to the complexity and size of the collected data.

The process of analyzing a digital image is usually divided into several steps, where the objects of interest are first identified, or 'segmented', followed by extraction of measurements and statistical analysis. This chapter starts from the basics of describing images as matrices of pixel intensities. Emphasis is thereafter put on image segmentation, which is often the most crucial and complicated step. A number of common mathematical models used in digital image processing of microscopy images from biomedical experiments are presented, followed by a brief description of large-scale image-based biomedical screening.

Keywords Image cytometry • Fluorescence microscopy • Cell segmentation • Image analysis

C. Wählby (✉)
Centre for Image Analysis, Division of Visual Information and Interaction, Department of Information Technology, Uppsala University, Uppsala, Sweden
e-mail: carolina@cb.uu.se

V. Zazzu et al. (eds.), *Mathematical Models in Biology*,
DOI 10.1007/978-3-319-23497-7_1

1 Pixels and Color Channels

A digital image is not continuous, but consists of discrete picture elements, or pixels. A typical fluorescencemicroscopy image is built up of multiple fluorescence channels, each representing a separate fluorescence stain, usually bound to DNA or an antibody probing a specific protein or subcellular structure. Figure 1 shows a fluorescence microscopy image where cell nuclei are stained with DAPI binding DNA, and red and green dots representing mRNA molecules (for details see [12]). Imagine that the goal of the analysis is to count the number of red and green dots per cell. The color image in Fig. 1 can be split into its constituent image channels, leading to one image representing the red, green and blue fluorescence respectively. If we take a closer look at the red channel, and zoom in on one of the dots, we can see that the image is built up of square picture elements, or 'pixels' for short, see Fig. 2. Each of these pixels is represented as a number in the computer, where a higher number means a brighter pixel, and the whole image can be thought of as a matrix of numbers. In a color image, the three image channels represent the

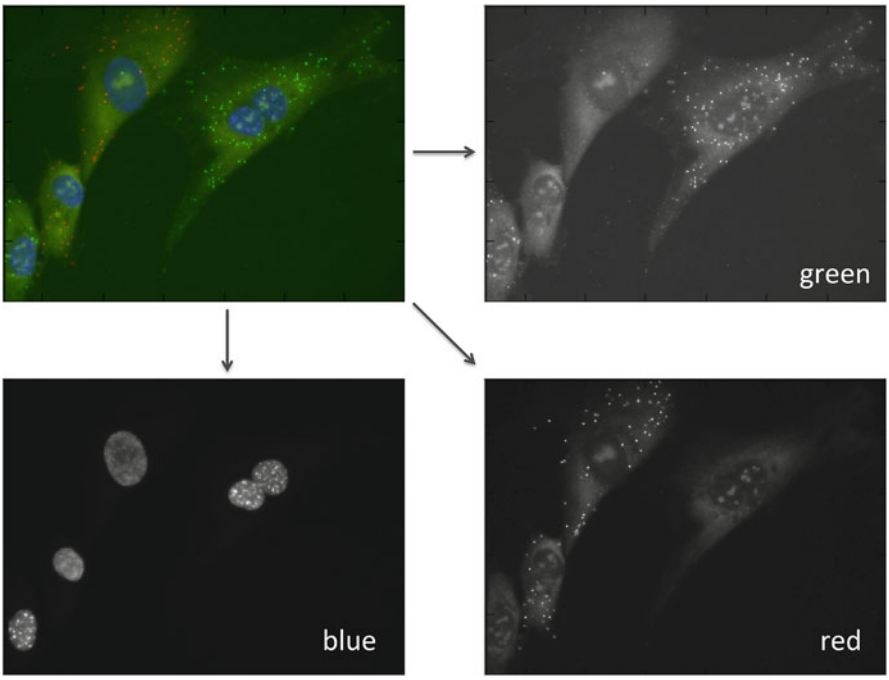

Fig. 1 Using three different filter sets, three different fluorescence labels were imaged using fluorescence microscopy. Top left is a composite image of all three image channels; cell nuclei are stained with DAPI binding DNA, and *red* and *green dots* represent mRNA molecules (for details see [12]). Due to autofluorescence and unspecific fluorophore binding, the cells' cytoplasms can be seen as a weak background staining in the *red* and *green* image channels