Computational Oncology

Alan T. Lefor*

Department of Surgery and the Jichi Medical Simulation Center, Jichi Medical University, Tochigi, Japan

*For reprints and all correspondence: Alan T. Lefor, Jichi Medical University, Yakushiji 3311-1 Shimotsuke City, Tochigi 329-0498, Japan. E-mail: alefor@jichi.ac.jp

Received February 7, 2011; accepted May 19, 2011

Oncology research has traditionally been conducted using techniques from the biological sciences. The new field of computational oncology has forged a new relationship between the physical sciences and oncology to further advance research. By applying physics and mathematics to oncologic problems, new insights will emerge into the pathogenesis and treatment of malignancies. One major area of investigation in computational oncology centers around the acquisition and analysis of data, using improved computing hardware and software. Large databases of cellular pathways are being analyzed to understand the interrelationship among complex biological processes. Computer-aided detection is being applied to the analysis of routine imaging data including mammography and chest imaging to improve the accuracy and detection rate for population screening. The second major area of investigation uses computers to construct sophisticated mathematical models of individual cancer cells as well as larger systems using partial differential equations. These models are further refined with clinically available information to more accurately reflect living systems. One of the major obstacles in the partnership between physical scientists and the oncology community is communications. Standard ways to convey information must be developed. Future progress in computational oncology will depend on close collaboration between clinicians and investigators to further the understanding of cancer using these new approaches.

Key words: oncology – physics – mathematics – computers

DEFINITION

Computational oncology is the application of techniques from the physical sciences (e.g. physics, mathematics and engineering) to problems in tumor biology.

HISTORY

Applying the basic scientific laws of physics and mathematics to improve the understanding of living organisms is not a new concept. In 1943, Erwin Schrödinger (Nobel Prize, Physics, 1933) delivered a series of three lectures at Trinity College (Dublin) to explain to the public how the laws of physics, and especially thermodynamics, may be used to better understand the human organism. The third and final lecture focused on the concept of entropy and how it relates to biological systems. ‘Schrödinger’s Paradox’ refers to the fact that although living systems must obey the laws of physics, which include increasing entropy (disorder), living systems seem to decrease in entropy, becoming more ordered. He concluded that the laws of physics were inadequate to describe biological systems because they are by their nature not closed systems, and they use external energy to circumvent the second law of thermodynamics.

In 1825, Benjamin Gompertz published a demographic model of mortality, which was widely used for life insurance cost calculations. The law is based on an exponential function, which initially has a slowly increasing element followed by one of rapid increase and then a slowly increasing element again. In 1954, Armitage and Doll (1) developed a two-stage mathematical model of carcinogenesis. The Gompertzian model of tumor growth was used by Laird (2)
in 1964 to model tumor growth kinetics and has been taught widely around the world in oncology education.

A NEW APPROACH TO ONCOLOGY RESEARCH

Despite significant advances in the understanding of tumor biology, improved patient outcomes in terms of cure rates and survival have not met some expectations. The past 50 years have seen major advances in scientific knowledge and technology. Increasing amounts of money are being spent to meet the clinical challenge of cancer. However, further progress is certainly needed.

This situation may be compared with the problems associated with software development, as described by Frederick Brooks in The Mythical Man-Month (3). As the lead developer of the IBM System/360 computers, Brooks had a unique perspective on a very large engineering project. Work was defined in terms of man-months to complete a segment of the project. Despite the addition of increasing numbers of people, delays mounted in this massive project. Simply adding more people and more money do not necessarily result in reduced time for development. In some ways, this is striking parallel to what has happened in oncology research over the last 50 years, although the analogy is not complete since scientific knowledge and progress in engineering are different enterprises.

In order to make progress in oncology, more investigators and more grant money have been added, but this may not lead us more quickly to more effective therapy. Further progress may require an entirely new and novel approach. Recent developments in computational oncology may herald such a new approach. While in the past, oncologists and oncology investigators saw themselves as entirely separate from mathematicians and physicists, the novel idea to bring these once-disparate groups together to work on a common problem is long overdue and has spurred the development of the field of computational oncology. A ‘think tank’ recently convened to identify opportunities and barriers for leveraging the physical sciences in oncology (4). Four major themes emerged from this conference including:

- A sound mathematical foundation and standards for communication are required to bring research in the physical sciences and oncology together.
- Advanced technologies such as nanotechnology and mathematical models are necessary to develop an understanding of how physical laws govern complex biological systems.
- Cancer must be viewed as an evolutionary process, and this may allow the development of new approaches across the spectrum of oncology.
- Information flow between cells and between cells and their environments must be better understood and may be facilitated by applying technology from the physical sciences.

Computational oncology must be placed in the context of the entire field of biology and medicine. The concept of systems biology is a very recent one, which has been advanced by Kitano (5). This approach to the understanding of biological processes requires an understanding of how a system is interconnected, in a structural and dynamic way (Fig. 1). This thought process leads to the notion that in a biological system, large numbers of functionally diverse elements interact selectively and non-linearly to result in coherent behaviors (6). It is the combination of elements that result in a specific behavior. The real power of mathematical modeling of such systems is in using mathematical tools to relate the multiple components of a complex system and further understand the behavior of the system (7). A cycle of research will begin with identification of a problem and the use of computational tools to model that problem. After analysis of these initial results, experiments can be designed and carried out based on the results of modeling and systems analysis (Fig. 1).

The desire to understand the underlying complexity of biological systems along with developments in molecular biology and recent improvements in computing power has led to the development of the field of computational medicine. Table 1 shows some of the areas where research in computational medicine is directly impacting the clinical practice of medicine in a wide range of specialties. Computational oncology is just one element in this rapidly expanding field and can be conceptualized as shown in Fig. 2. The organizational elements shown in Fig. 2 are not intended to be comprehensive, but rather illustrative of the diversity of research now underway. The divisions between areas of research are not intended to be sharp as collaboration is very important to make further progress.

Computational oncology has evolved in a very short time and is developing along two parallel, but separate pathways. The first path is the use of analytic techniques from the physical sciences to evaluate the massive quantities of data that
Table 1. Clinical applications of computational medicine

<table>
<thead>
<tr>
<th>Clinical field</th>
<th>Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oncology</td>
<td>Cancer database analysis to establish trends in diseases and identify gene targets for new therapeutic approaches</td>
</tr>
<tr>
<td>Oncology</td>
<td>Evaluation of genomics data for the design of cancer clinical trials</td>
</tr>
<tr>
<td>Oncology surgery</td>
<td>Treatment planning using imaging data for the resection of tumors</td>
</tr>
<tr>
<td>Oncology</td>
<td>Tumor growth models to optimize treatment timing and develop personalized therapy</td>
</tr>
<tr>
<td>Oncology</td>
<td>Development of new anti-cancer drugs by mass screening with computers</td>
</tr>
<tr>
<td>Oncology</td>
<td>Analysis of data sets to predict tumor recurrence</td>
</tr>
<tr>
<td>Radiation oncology</td>
<td>Computerized integration of imaging data to design optimal treatment plans</td>
</tr>
<tr>
<td>Breast oncology</td>
<td>Breast modeling—imaging for breast conserving surgery planning</td>
</tr>
<tr>
<td>Breast oncology–radiology</td>
<td>Computerized analysis of routine imaging studies such as mass screening of mammograms</td>
</tr>
<tr>
<td>Surgery</td>
<td>Control of surgical robots in neurosurgery and general surgery</td>
</tr>
<tr>
<td>Surgical training</td>
<td>Laparoscopic training and evaluation</td>
</tr>
<tr>
<td>Vascular surgery</td>
<td>Analysis of stent function, modeling of vascular grafts</td>
</tr>
<tr>
<td>Interventional radiology</td>
<td>Robotic insertion of needles for image-controlled biopsy procedures</td>
</tr>
<tr>
<td>Cardiology</td>
<td>Image analysis and refinement for CT angiography of the heart</td>
</tr>
<tr>
<td>Cardiology</td>
<td>Systems biology analysis of the heart in congenital heart disease</td>
</tr>
<tr>
<td>Pathology</td>
<td>Automatic screening of cytology specimens for malignant cells</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>Computational models of Schizophrenia mechanisms</td>
</tr>
<tr>
<td>Ophthalmology</td>
<td>Studies of eye trauma models to evaluate mechanisms of injury to the eye</td>
</tr>
</tbody>
</table>

Figure 2. Computational oncology is part of the rapidly growing field of computational medicine, and may be thought of as having two major areas of investigation including data analysis and tumor models. The divisions between these areas should be blurred to maximize collaborative research.

are now available, especially from genomic investigations (data analysis). The second path is the use of techniques from the physical sciences to create mathematical computer-based models of tumor growth, metastatic potential, vascular supply etc. in an effort to understand the physical nature of tumor growth and ultimately to define optimal therapeutic
strategies (tumor models). In this review, both of these paths will be explored, as well as suggestions for how the oncology community can contribute to this new and exciting effort.

**COMPUTATIONAL ONCOLOGY: DATA ANALYSIS**

Computers have been used for data analysis from their very inception. The ability of computers to analyze larger and larger volumes of data is a result of marvelous technologic developments including faster processors and vast amounts of cheap memory. The study of systems biology requires the modeling of vastly complex components, which interact with each other on many levels. These models are crucial to the understanding of complex biological systems and are referred to as *in silico* models (8). The dysfunction of any component of the network can lead to diseases such as cancer, and the use of computational oncology techniques has allowed the development of sophisticated models to deepen our understanding of the origins of malignancies. Much of this understanding is a result of using computers to analyze vast quantities of data that are collected in various databases. As a first step, it is essential for investigators to have a common language, in order to facilitate communication among investigators. While this may seem self-evident, it was identified as a major issue in the development of collaborative research between oncologists and those in the physical sciences (4). Efforts to accomplish this goal are exemplified by the Systems Biology Markup Language, developed in an effort to form a *de facto* standard and open software platform (6).

**CELLULAR PATHWAY DATABASES**

The development of a computational model of cellular processes from available databases has been outlined by Daskalaki et al. (8). There are three basic components including pathway databases, annotation tools and simulation tools. Proteins that have been identified as possibly important are searched in pathway databases, and a molecular network is then constructed using information from the Pathway Databases Reactome and Consensus PathDB. The model thus constructed is then analyzed with a modeling tool. There are multiple tools available to perform each of these functions, including the Systems Biology Workbench, allowing the combining of data from heterogeneous components.

**MICROARRAY ANALYSIS**

The first oncogene was discovered in 1977, and since then, research in this field has exploded with a wide variety of techniques for the identification of cellular genetic aberrations (9). The analysis of data from microarrays is only possible because of the use of computers. Microarrays have been used increasingly in many areas of cancer research and represent a major advance in biomedical research. The recent completion of human genome sequencing has advanced this field at an incredibly rapid rate. The techniques thus developed will hopefully yield important results about tumorigenesis, more accurate and rapid diagnosis, more comprehensive prognosis and ultimately improve therapeutic interventions to benefit patients. All of these developments are dependent on computer technology and the techniques of computational oncology.

Simultaneous monitoring of thousands of gene expressions per sample is now possible using microarray technology (10). More recently, mRNA expression data have been correlated with survival, allowing for the prediction of patient survival in other patients. Tissue samples from patients with B-cell lymphoma and breast cancer were studied for the expression of 5622 genes using a novel proportional hazards model, and patient survival probabilities calculated (10). The models thus generated were then used to predict survival probabilities.

In microarray genomics, expression data from normal tissue and from malignancies are compared to identify differentially expressed genes that may prove important in follow-up validation studies (11). The initial identification of these genes is thus a fundamental issue in genomics and one that has been facilitated by computational oncology techniques. One of the problems with gene expression data is that it exists in multiple dimensions and is thus complex. Mathematical techniques have been developed to reduce the dimensionality of these vast data, and the partial least squares method is of particular importance in cancer genomics (11). Another problem in the use of microarray data is that there are so many genes being studied at once and that the identification of specific marker genes becomes a significant statistical problem. The number of genes is often greater than the number of arrays, which makes the problem mathematically complex (12). One approach to this problem has been the selection of relevant subsets of genes, which improves the prediction performance and simplifies the mathematical problems. This has been approached using machine learning techniques to select these subsets, including recursive feature elimination, nearest shrunken centroids and random forests (12).

Other techniques have also been developed to address the complex problems associated with the identification of significant genes. One of these is referred to as the ‘wrapper method’ which utilizes decision tree analysis, called the Classification and Regression Tree (CART) technique (12,13). Classification algorithms to identify genes that are efficient for classification were applied to large public databases for leukemia, colon cancer and prostate cancer. This technique led to the identification of genes with biologically relevant activity. These investigations have also emphasized the importance of including clinical variables in the studies.

Although data from microarray analysis have become very important, combining these data with clinical data leads to more useful computational models. Bayesian networks are
used as decision support models and are of value because they specifically model the uncertainty in data, by combining both probability theory and graph theory (14). The prediction of prognosis in patients with lymph node-negative breast cancer was modeled by Gevaert et al. (14) using a Bayesian network. This model was developed using data from an available database and looked at 25,000 expression values per patient. Clinical and microarray data were combined in the Bayesian network with three levels of integration. The integrated use of clinical and microarray data outperformed indices that are based on clinical data alone, and demonstrated the value of Bayesian networks in computational oncology.

In another example of clinical applications of this technology, investigators constructed a microarray from over 2,500 prostate cancer specimens and included in their analysis histopathological and clinical follow-up data (15). Gene expression and copy number were analyzed for 16 candidate markers for their ability to predict tumor progression and patient prognosis. The data from this microarray analysis were used to extend established clinical prediction tools, which had been based on non-molecular data only. Using this combined approach, a marker was identified that increased the accuracy of the clinical nomogram, and fulfilled the criteria of a novel marker for prostate cancer (ANXA3). By integrating large-scale clinical and molecular data, new markers may be identified in the future.

New approaches to microarray technology are constantly being developed. One of the most promising of these is the field of microfluidics, which has created the ‘Lab-on-a-chip’ technology (16). These techniques allow on-chip cytometry and microsorting techniques. This exciting technology will likely have wide application in the future of pharmacologic screening, drug discovery, in vitro oncology models and clinical oncology.

**CELLULAR IMAGING AND DRUG DEVELOPMENT**

High-throughput screening (HTS) is an integrated technology utilizing images of living cells as the basic unit to produce information on a wide range of responses including gene manipulations, drug responses and others in both normal and abnormal cells (17). HTS has been employed since its advent in the mid-1990s, for both basic and applied research. The ability to acquire and analyze imaging data rapidly is an obvious application of computing technology. This is being commonly done with automated fluorescence microscopy.

These techniques are highly dependent on fast computing processors as thousands of cells are analyzed. The software must be able to measure cancer-related changes in the cellular phenotype that results from gene manipulation, and includes genome instability, cell proliferation, apoptosis, angiogenesis, invasion and metastasis (18). HTS hardware has resulted from continued refinement and easily fits on a desktop as a single integrated unit (e.g. ArrayScan, Thermo Scientific, Waltham, MA, USA). The future of HTS will probably see further improvements in speed of throughput, but also in a greater focus on assays of increased physiologic and oncologic relevance (19).

The automated study of cell cycle progression is an important part of modern anti-cancer drug development (20). Image data are acquired and computational techniques used to examine segmentation, feature extraction, classification and tracking of individual cells (17). The use of fluorescent probes and image analysis has allowed a detailed analysis of the process of mitosis, and thus the investigation of the effects of many potential anti-cancer compounds in a short period of time. The sophisticated software now being used allows the extraction of features to discriminate among the different cell phases. Models are then created mathematically, and the features of new cells are compared with the models to allow classification. These models are updated using data from various sources, and then improved in their ability to classify cells according to the phase of the cell cycle.

HTS technology can rapidly assess the cellular effects of possible anti-cancer compounds. In a study of apoptosis, the effects of NSC95397, brefeldin A and bortezomib on several cell lines were studied. After incubation of the cell lines with the compounds, fluorescein-tagged probes were bound to the cells and quantitative measurements made using automated HTS technology for image capture and analysis (21). This study showed that all three drugs led to an increase in caspase-3 activation with modest changes in nuclear morphology. This study is an example of the cellular studies possible with HTS technology in the future of oncology.

The computer modeling of cell morphology has also examined the influence of various extrinsic and intrinsic factors (22). The models thus created were used to create collections of multicellular morphologies arising from the systematic modification of model parameters. By examining the effects of changing these parameters, values leading to the formation of robust epithelial structures were identified. Other values were identified that led to the development of abnormal tumor-like morphologies. In this study, cell morphology was used to adapt in silico models, examining the effects of a limited number of parameters. Thus, cellular imaging and morphology data can be combined with cellular models to create more accurate models.

**ANALYSIS OF ROUTINE IMAGING DATA**

Advances in computational power and algorithms have naturally led to an expanded role for the computer analysis of routine imaging data. The use of automated techniques to assist in the interpretation of radiographic imaging has become particularly important with the widespread screening programs that are now in place in many countries around the world, particularly for common diseases such as breast cancer and lung cancer.

The use of computer-aided detection (CAD) for breast cancer has increased significantly in the last few years for a
number of reasons. First, screening mammography for breast cancer has become more widespread. In addition, it has been shown that a review of mammograms by a second radiologist increases cancer detection rates; yet, this is not always possible because of increased cost as well as a shortage of radiologists (23). Increased detection rates for breast cancer are important because survival is improved when cancers are detected early.

CAD reviews mammograms with software to alert the radiologist to findings potentially associated with breast cancer such as microcalcifications and focal densities. In a study of CAD of breast cancer, 9520 consecutive mammograms were reviewed (23). The positive predictive value of biopsy recommendation, biopsy rate and recall rate were compared before and after the introduction of CAD. Size, stage and histology of cancers were also compared. This study found a significantly increased rate of detection for ductal carcinoma in situ (DCIS), with 14.2% more lesions identified by CAD. The additional cancers detected by CAD were smaller than without CAD. The screening recall rate increased from 6.2 to 7.8% after CAD, with a decrease in biopsy rate. This study demonstrates the benefit of CAD in evaluation of screening mammograms.

The algorithms used to identify cancer on mammograms are being developed and improved rapidly, using a variety of techniques in statistics and artificial intelligence. One such approach was outlined in a study by Rajendra-Acharya et al. (24), which starts with conversion of the image to a gray-scale form. The quality of the image is improved, and a uniform histogram is obtained. The image is then binarized with a suitable threshold, and the resulting image examined for three features including area, homogeneity and microcalcification. These three features allow classification of the image into normal, benign or cancer. Using a database of 360 images, the software was ‘trained’ to recognize the mammograms with cancer. The resulting accuracy was 88% in the study. Clearly, further improvement is needed, but these three features appear to be an important approach to develop software for mammogram analysis.

Another technique was described by Verma (25) and uses a clustering algorithm to cluster feature data for both benign and malignant classes of mammogram. Strong clusters are selected and used for final classification as benign or malignant. Features associated with benign and malignant images are contained in a neural network, using a multicluster class-based approach. This novel approach has a 96% classification accuracy using test data, which is much higher than results with a standard classifier system which uses a single cluster per class.

Lung cancer is also increasing in incidence and the objective of many screening programs. The detection of lung cancer at an early stage is also associated with an improved prognosis. Taken together, these facts support the importance for the development of improved methods for screening, including CAD techniques. CAD systems have been developed for the identification of pulmonary nodules on computer tomographic (CT) scan data. In a study by Awai et al. (26), 82 cases were evaluated. The CAD system identified 62 pulmonary nodules and missed 16 nodules for a true-positive rate of 80%. The study evaluated 3556 sections, with nodules of a mean size of 0.81 cm for the 16 missed nodules and 0.91 for the 62 correctly identified nodules. When the CAD software was used along with radiologist interpretations, significant improvement in lung nodule detection rates was reported. Although further refinement of the technique is needed, the value of this approach is clear. New approaches to imaging are increasing the challenges to software development including the use of volumetric medical imaging (27). Another limitation in the development of CAD approaches for lung lesions is the variation in training and testing data sets. The Lung Image Database Consortium (LIDC) was recently developed to overcome this limitation, by providing an extensive library of chest CT scans online. The LIDC database can be used for education, training and the development of new approaches to CAD of lung lesions (27). The relatively recent use of magnetic resonance (MR) imaging in the evaluation of lung nodules means that CADS approaches for MR imaging of lung lesions are not yet reported. The development of such approaches may be facilitated by the availability of an MR database such as the LIDC.

CANCER PROTEOMICS, GENOMICS AND DATABASES

The Human Genome Project has led to the development of the new fields of genomics, proteomics and now cellomics, which are evolving at a rapid rate. Proteomics may be considered as a mass-screening approach to molecular biology, to document the overall distribution of proteins in cells, characterize proteins of interest and ultimately to elucidate their functional role (28). Genomics examines the functionality of specific genes, their relation to diseases and their participation in biological processes. Cellomics looks at the phenotypic changes induced by alterations in various genes. Each of these areas is highly dependent on computation for data collection and analysis.

Although current medicine employs molecular markers on the scale of a single gene, the future will see utilization of thousands of markers simultaneously. These molecular-oriented diagnostic techniques will be linked with the prediction and prevention of disease on a molecular level (29). These developments are directly dependent on the future of computational oncology to combine large-scale and genomewide data with biostatistical and bioinformatics analyses of model systems. Such a complex analytical process is the basis of a new scientific paradigm, known as ‘integrative genomics’.

The identification of proteomic biomarkers in the blood is a complex process involving blood sample preparation, biomarker candidate discovery using analytical algorithms, and validation and clinical application of proteomic biomarkers (30). Although tumor markers are part of the clinical routine

Downloaded from https://academic.oup.com/jjco/article-abstract/41/8/937/816099 by guest on 17 March 2019
used commonly throughout the world, they lack sensitivity and specificity thus precluding their use in population-based screening. New techniques are needed to effectively use tumor markers. Genomics and proteomics are promising approaches to this complex problem. The most prominent methods in use today include serological analyses of recombinant DNA expression libraries, gel electrophoresis, mass spectrometry and protein/DNA microarrays (31). In addition, a direct computational approach allows detection of candidate tumor marker genes in publicly available data sets.

One of the most promising approaches to the proteomic identification of marker genes has been the use of surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) mass spectrometry. The use of proteomic patterns in serum to diagnose ovarian cancer has been reported to result in an excellent sensitivity and specificity (32). Further studies are needed to validate these results, but the approach is very promising. The SELDI-TOF method is associated with some difficulties however, and further refinement of the technique is needed (33). It does suggest, however, that the proteomics approach, supported by computational techniques, has promise to identify clinically useful tumor markers.

One of the problems associated with improved proteomic approaches is the need for improved computational algorithms. There are two principal problems being encountered in the analysis of the mass spectroscopy data including the extraction of appropriate features that represent the identities of different classes as well as how to handle the huge quantities of data generated (34). Efficient analysis of the data requires non-linear algorithms, which are a new approach. One such technique is referred to as ‘evolutionary computing’ which is highly efficient in searching in high-dimensional space, especially for non-linear optimization problems (28). This approach has been applied to the analysis of SELDI-TOF data. Evolutionary computation is a system that mimics the adaptation and evolution of each individual in its environment, analogous to Darwin’s theory of evolution. This approach encompasses a wide range of techniques including genetic algorithms, genetic programming, evolutionary programming and evolutionary strategy. As an example, genetic programming consists of trees of nodes based on decision analysis. The good trees are then copied from one generation to the next, functioning as evolution, based on classification of the results of analyzing the tree. Bioinformatics clustering analysis has also been used to efficiently and accurately analyze proteomic data (35). This approach has also been applied to SELDI-TOF data and is based on clustering algorithms derived from Bayesian analysis. One of the problems with such data is that there is a great deal of ‘noise’ in the results. The use of clustering analysis has resulted in removal of up to 99% of the noise, and thus enabling the more efficient identification of significant proteins for further analysis.

SELDI mass spectrometry has been applied to specific clinical problems. In a study of lung tumor tissue, SELDI revealed three peaks at the 17–23 kDa mass range from tumor cells, which were markedly increased compared with normal tissues (36). The identification of this protein signature, which is highly reproducible in malignant lung tissue, may lead to a screening technique that can be used to study populations at risk for lung cancer and monitor the response to chemopreventive agents. In another study, SELDI-TOF was used to profile the serum from 36 patients with renal cell carcinoma (37). The sera were classified using an artificial intelligence algorithm, and five proteins identified which could correctly separate serum from patients with renal cell carcinoma from serum from health patients. The combination of the SELDI-TOF data with computational algorithms may be a powerful tool in the future for screening and diagnosis.

In addition to the identification of biologically significant proteins, investigators are aware of the interactions of various proteins. Investigations of these interactions are highly dependent on computational algorithms. Experiments that document such interactions are commonly reported through the International Molecular Exchange (IMEX) consortium and must be compliant with various guidelines (38). As of September 2009, the IntAct database contained more than 200 000 binary interaction evidences. This database is available for query by investigators with a text-search facility as well as allowing the use of the Molecular Interaction Query Language. The growth of IntAct as more data regarding molecular interactions will play a major role in the future of systems biology.

The ability to extract meaningful data from ever-expanding databases is an important area of development in computational oncology. Specifically, the relationship between genes and cancer is being documented by data mining from large databases. One example of this is the work of Zhu et al. (39), who developed a probabilistic model for mining data from the Online Mendelian Inheritance in Man (OMIM) and Medline databases. They developed a system using a mixed aspect model, and refined the probability parameters using a learning algorithm. They focused on genes in cancer gene co-occurrence pairs from the OMIM and they further analyzed these positive pairs. A total of 3118 gene–gene co-occurrence pairs were identified and then reduced the size of the data set by looking at co-occurrences. Different mathematical models were then constructed. By using this technique, they were able to identify previously unknown genes that are associated with major cancers and are worthy of further study. In the future, gene co-occurrence data may be combined with microarray expression data to further characterize these cancer associated genes. This technique demonstrates the power of analyzing data in vast databases.

Computational techniques are being applied not only to the identification and study of individual cancer-related genes, but also to the construction of models to study gene regulatory networks, involving many genes. However, the computing time required for such models can be formidable. One of the common modeling tools used to study these
networks is called MEDUSA, which needs more than 4 weeks of computing time to model and analyze 7000 genes using 1000 iterations (40). Advances in algorithm design and implementation led to the development of FastMEDUSA, which models gene sets about 40 times faster than the original implementation. Although MEDUSA takes 12 days to analyze the human glioma data set in 400 iterations, FastMEDUSA analyzes the same data set in ~6h. Improved computing algorithms are an essential part of the future of computational oncology.

COMPUTATIONAL ONCOLOGY: TUMOR MODELS

Using mathematical models to understand the behavior of cancer has generally proceeded in two directions, including descriptive models and mechanistic models (7). Descriptive models may seem most intuitive to clinical oncologists and examine the gross characteristics of a tumor such as its size, growth pattern and overall dynamics. Mechanistic models focus on the various processes that lead to tumor growth in an effort to understand the relative contributions of the various components to overall tumor behavior. Both of these approaches are important for further understanding of tumor biology. The use of data from in vitro experiments or from animal experiments to develop and improve these models has led to improved accuracy of the resulting model.

DESCRIPTIVE MODELS

Although experimental research using computational methods is important, synergistic development by the addition of clinical information in computational oncology is extremely important. Descriptive models of tumors (Fig. 1) are especially useful when the model is refined by patient data, enabling the development of patient-specific therapy based on their own tumor growth pattern. In a study of 32 patients with gliomas, investigators found that dynamic insight from pre-treatment imaging may be quantitatively useful in predicting the survival of individual patients (41). A mathematical model was first developed, using a reaction–diffusion partial differential equation that equated the rate of change of glioma cell density to the net dispersal of glioma cells plus the proliferation. This model allowed quantification of the extent of the tumor as well as predicting radial expansion on routine imaging. The net result of combining this model with patient imaging data is an understanding of the glioma phenotype for each patient yielding patient-specific prognostic information. Using two sets of imaging data allowed measurement of the velocity of radial expansion of the tumor, and then using the model, net rates of dispersal and proliferation were calculated. This technique is the use of an ‘untreated virtual control’ group. The authors concluded that their model using the parameters of net proliferation and invasion rates was significantly associated with prognosis. Further studies of this approach, perhaps in other patient populations, will hopefully lead to clinically useful results.

A novel type of descriptive model has been used to simulate the behavior of DCIS which allows not only a physically accurate description of the system, but is also refined by patient data (Fig. 2), and is referred to as agent-based modeling (42). Each cell is considered an object subject to classical laws of physics, particularly the conservation of momentum. The cells are described by their mass, radius, volume, solid volume, position and velocity. Phenotypic properties such as cell state, calcification, surface receptors, proliferation and apoptosis were also built into the model. Cell motion was modeled based on the balance of forces and included mathematical descriptions of cell velocity, cell–cell adhesion, cell-basement membrane adhesion, cell–cell repulsion and other forces. Ductal geometry was also modeled. Initial values were used, and then the model refined from patient information was obtained from histologic sections. This model may enable accurate predictions of the excision volume required to completely excise DCIS lesions surgically.

MECHANISTIC MODELS

DISCRETE CELL MODELS

Mechanistic models of tumor growth are usually divided into two groups, including discrete cell-based models and continuum models (Fig. 2) (43). In discrete cell models, biophysical rules are used to model individual cells. This approach is considered especially useful for studying carcinogenesis, genetic instability and interactions of single cells with each other or with the environment. However, as one tries to model more cells, the computational power required increases at a rapid rate.

CONTINUOUS MODELS

These models use mathematical equations to model the tumor as a homogeneous population. In a study to simulate solid tumor growth, Cristini et al. (44) used a complex non-linear simulation to examine tumor cell invasion and branching. In this study, partial differential equations were used to model the non-linear effects of cell–cell adhesion and taxis including chemical and molecular species. Due to the complexity of the equations used, numerical methods were developed to solve the equations. This model accurately predicted that nutrient deprivation can importantly contribute to increase the invasive behavior of a tumor. They also showed that diffusional instability is exacerbated by nutrient gradients.

Complex tumor models are made possible in part by the improvements in computing power that are now widely available. A diffuse interface model of tumor growth was developed by Wise et al. (45). This multispecies model accounts for a variety of mechanical interactions among
cells, including volume fractions of water, tumor cells, host cells, densities, extracellular fluid pressure, cell-to-cell pressure and component velocities. Detailed models were developed and solved using complex numerical methods. The models use complex high-order non-linear partial differential equations which can only be solved using numerical techniques. These methods led to the development of a diffuse interface continuum model of multispecies tumor growth. Although the initial work looked at a two-dimensional model, extension to a three-dimensional model was then undertaken (46,47). This model showed that morphological instability leads to invasion of tumor cells and suggests that combining such models with patient tumor-specific parameters may lead to improved patient care.

**Hybrid Models**

The hybrid discrete continuum (HDC) model has led to some very interesting conclusions which challenge the common paradigm of tumor behavior. This model examines the interface between the cancer cell and its microenvironment (7). Each cell is modeled as a point on a latticework that represents the microenvironment of the tissue, and the traits of the cell are modeled in regard to proliferation, migration, adhesion and nutrient consumption. Each of the traits is quantified by coefficients which can change as the cell changes. As traits of individual cells are altered by changing these coefficients, the system behaves as an evolutionary process. This model has led to the novel conclusion that cells become invasive collectively, whereas individual cells may not show such behavior. The HDC model suggests that the tumor cell microenvironment has a selective effect on invading cancer cells, rather than serving as a supporting infrastructure. Further experiments with living cells will be needed to validate these novel concepts, but the fact that these results are from mathematical models shows the power of tumor cell modeling.

The interface between mathematical modeling and traditional experimental approaches remains an important area for collaboration. Cristini et al. (44) developed a non-linear model of tumor growth that used a reaction–diffusion mathematical model. This model predicted that mass growth and shape of in vitro spheroids depend on four microphysical variables, including cell mitosis rate, diffusion length and two dynamics-controlling dimensionless variables. Using this model, the growth of tumor spheroids was modeled. Then, using an in vitro model of tumor spheroid growth using glioma cells, computational results and experimental in vitro observations of morphology were in close agreement. The authors concluded that the tumor morphology observed in vitro is the result of shape instability driven by diffusion gradients in the tumor microenvironment as predicted by the mathematical model. This model has improved our understanding of the contribution of the microenvironment to tumor morphology, and further experiments are in progress to examine this phenomenon in vivo.

The results described for various descriptive and mechanistic models represent early, yet very important studies in tumor cell modeling. The importance of discrete cell models as well as continuum models in the major category of mechanistic models is demonstrated. These results show that it is possible to utilize descriptive models modified by patient-specific data to further the understanding of the clinical behavior of tumors. This approach could lead to highly individualized treatment regimens based on the growth dynamics of a patient’s own tumor. Improvements in computing power as well as the further refinement of mathematical models should result in a bright future for this discipline.

**THE ROLE OF THE CLINICAL ONCOLOGIST IN COMPUTATIONAL ONCOLOGY**

Although many of the activities in the field of computational oncology have been going on long before the field of computational oncology was considered as a separate entity, the collection of such activities under a single ‘umbrella’ will be useful to bring people together who have common interests and advance the field more rapidly. The concept of bringing together investigators with seemingly disparate interests such as physics, mathematics and oncology may seem strange at first, but it represents an important step to advance the knowledge and understanding of malignancies. More interdisciplinary programs must be developed in the future.

One of the most difficult problems in advancing the field of computational oncology will be to overcome the difficulties in communication among practitioners in the various fields. The development of common terminology and common software is essential to advance the field. The clinical oncologist, who may not be actively engaged in research in the field, is an essential participant. Clinical oncologists can be of enormous assistance in a number of ways. First, those in the physical sciences who are engaged in this field need guidance to understand the relative clinical importance of problems that they wish to solve. Some physical scientists may not have the clinical insight to provide such guidance and therefore collaborative approaches are essential. The clinical oncologist can also make significant contributions to computational oncology by providing data and specimens. Advances in computational oncology will depend on using patient information and tissue to modify and refine the computational approaches developed. This information must be willingly provided by clinicians, who have an essential role in the future of computational oncology.

The application of techniques from the physical sciences to problems in oncology comes at a time of significant advances in the application of computing techniques to problems in physical sciences. There is an extraordinary parallel between what is going on in the field of numerical
relativity and computational oncology. In numerical relativity, computers are being used to solve complex differential equations in an attempt to further understand the theory of relativity. This has only become recently possible with the development of supercomputers and new computational algorithms. The situation in the development of complex tumor models using systems of partial differential equations is remarkably similar, and we can only hope that the future of this exciting new field will ultimately provide a benefit to our patients from the synergistic activity between investigators in the physical sciences and those in oncology.

Conflict of interest statement

None declared.

References


