Close Encounters of the Collaborative Kind

Michael B. Mayhew, Xin Guo, Steven B. Haase, and Alexander J. Hartemink, *Duke University*

The participants in a collaborative interdisciplinary project found that developing a shared, project-specific communication style helped them overcome cultural barriers, understand the nuances of each other's work, and enhance the accuracy, interpretability, and utility of their models.

Interdisciplinary research requires people with different perspectives to work together to share expertise and bridge gaps in the understanding of a particular problem.[1] This is certainly true in systems biology, a new and vibrant field that brings together researchers from a wide range of computational and experimental biological disciplines.[2]

Systems biology aims to quantitatively describe the holistic properties and behaviors of entire cells, organs, or populations, rather than simply the properties of their individual parts. Systems biology has helped bring about a paradigm shift in the kinds of biological questions posed in the laboratory. The field also offers a proving ground for the development of stronger, closer relationships between biological and computational researchers, resulting in better analysis, modeling, and, ultimately, understanding of the biological system under study.

Our intent here is not to describe a specific scientific result, but rather to provide insights into the development of an interdisciplinary collaboration that, for us, has been effective in generating scientific results. As such, we only loosely outline our results to provide a backdrop for a discussion of our collaboration, and most details are omitted to focus on the collaboration itself. That said, we do provide a summary introduction to our results, and interested readers are encouraged to read our published papers for further details regarding our science.

To distill some insights and highlight some potential pitfalls in establishing an effective interdisciplinary collaboration, we share our personal experiences with modeling cell division in the budding or brewer's yeast, *Saccharomyces cerevisiae*. Although we faced a variety of computational issues along the way, the collaboration's challenges extended far beyond writing code. The painstaking work of finding and developing a common language to address our problems together turned out to be a necessary goal of the collaboration, and not simply a by-product. From the outset of the collaboration, we had to commit to overcoming cultural barriers and false assumptions.

This commitment to constant and effective communication allowed everyone involved in the project to better appreciate what other members contributed, develop better intuition on how to move the project forward, and add to the common language that facilitated the work.

CULTURAL DIFFERENCES

An effective collaboration needs to be a team effort directed toward a set of common goals, but scientific partnerships often look more like an uneasy yoking of two unequal parties with distinctly different goals.

In one version of such a scenario, computational researchers might start a project by reading a few biological articles and downloading publicly available data before embarking on their primary goal: the data-driven devel-

 Published by the IEEE Computer Society

opment of models or analysis tools. Once complete, the researchers might submit their work to a computational conference to highlight the utility of their approach—often demonstrated by something akin to a receiver operating characteristic curve. Once they have proven their method in this fashion, the computationalists might try to convince biologists to use their new model or analysis tool, or to validate its predictions.

In a somewhat more collaborative variation on the theme, computational researchers might start by expressing a desire to model or analyze some biologists' experimental data. Should the biologists consent, the computational researchers might then receive the data—along with an overview of the biological problem—so that they can carry out their data-driven analysis or modeling. The biologists might not be included in the process of developing the methods or algorithms; rather, this would be treated as a form of sophisticated black art, not to be revealed to the uninitiated. If the computational model's assumptions are consistent with biological reality, then so much the better, but this sort of correspondence does not outweigh, say, the novelty and elegance of the methods themselves.

In a third scenario, biological researchers might approach computational researchers, looking for partners to help them analyze their exciting new data. The experiments have already been designed and carried out, the data already exists, and the process of finding someone to help make sense of all the data begins only at this late stage. The computationalists are left to crunch some numbers and hand the data back to the biologists in a form that will lead to a figure in the eventual paper.

Clearly, these are not models of effective collaboration, and yet some variant of each of these scenarios crops up again and again. Why? At least part of the problem can be attributed to cultural differences between computationally trained and experimentally trained scientists.

DIFFERENT PERSPECTIVES

Cultural differences are not new, and have often proven to be stumbling blocks in the formation of new collaborations between biological and computational researchers.

Naturally, biological and computational researchers are trained differently. Biologists are often driven by questions or hypotheses regarding how or why certain biological processes occur. In contrast, computational researchers tend to focus on the principles behind the modeling or analysis of a particular dataset. Biologists might not care too much about the performance of some algorithm, while computational researchers might not be concerned about the details of experimental data acquisition.

For example, suppose computational researchers approach a biologist with a particular model they have developed. The model might be sound, well thought-out, and mathematically elegant—work that would be considered excellent among computational colleagues. However, if that model does not generate new insights or results related to the questions or hypotheses in which the biologist is interested, the computational researchers will have no end of difficulty convincing the biologist to participate in a collaboration. Biologists will have minimal interest in the success of the computational project if their questions cannot be addressed.

Winning the confidence of biologists can be an especially difficult task for the computational researcher who simply analyzes or models publicly available data without interacting with members of a biological community. Suppose again that a computational researcher has a model built solely from publicly available data. As with any model, some predictions will be incorrect. However, because of the manner in which the model was constructed, some of these false predictions will flatly contradict known biological facts. For example, a protein might be predicted to reside in the nucleus and yet is known from previous research—or from unpublished experimental data known only to biologists—to reside elsewhere. These kinds of false predictions will raise doubts in the minds of potential biological collaborators about the utility of the computational work and could even scupper the partnership before it starts.

Our collaboration could not have gotten off the ground if not for an initial series of deeply probing meetings.

COMMUNICATION LAYS THE FOUNDATION

Our collaboration could not have gotten off the ground if not for an initial series of deeply probing meetings. It cannot be overstated how important these early interactions were in grounding us in the realities of problem-solving from each other's perspectives, establishing common goals for the project, and overcoming language barriers.

When we first met, the computational contingent already had ideas about how to model transcription during yeast cell division and how to generate a very-high-resolution view of the process. However, some of the ideas related to acquiring new data were impractical. Also, the ideas we had about the process of cell division from textbooks or literature reviews often had to be revised or supplemented by more current—and often unpublished—experiments and insights known only to our biological colleagues.

In initiating the collaboration, we appealed to the goals of our biological collaborators, motivating each phase of the project with a biological question or hypothesis.

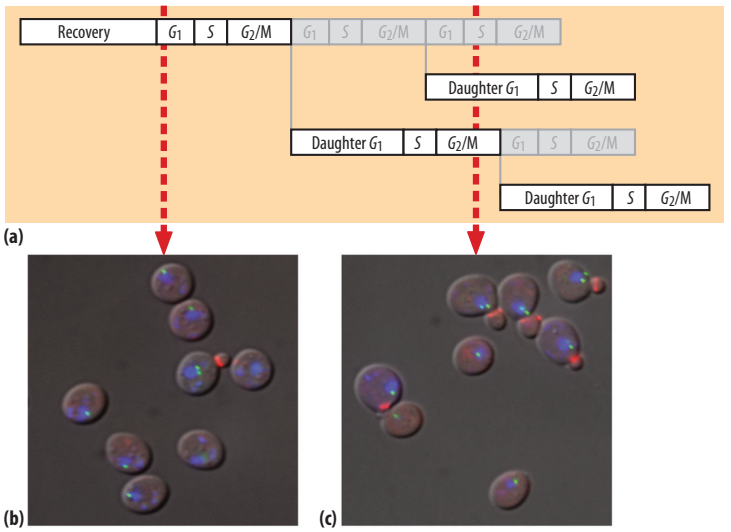

Figure 1. Branching process construction of the CLOCCS model captures loss of synchrony in cell populations. (a) Diagram depicting the branching process underlying the CLOCCS model. The cell population comprises different subpopulations. The initial population undergoes a recovery period after release from synchronization before entering the cell division cycle. After division, a new branch is created along which a subpopulation of daughter cells progresses. (b) At early points in the time-course experiment, when synchrony is still fairly good, the population is largely homogeneous in terms of cell-cycle progression. (c) At later time points, with the loss of synchrony, the population is a more complex mix of different subpopulations at different stages of the cell cycle. Red dashed lines indicate points in the time-course experiment.

Framing the question as a computational task and developing appropriate methods began only after establishing the biological question itself. At the same time, thinking carefully about what methods might be applicable helped us determine the kind of data required and design the appropriate experiments for acquiring that data.

This approach of appealing to the biologist's goals and using biological questions to motivate method development has worked well for us in practice, but it is by no means the only way computational researchers might begin collaborating with biologists. In other situations, computational researchers might already have developed a model or analysis tool before approaching biologists. In such a scenario, computationalists and biologists can fruitfully discuss what exactly the method is capable of revealing, and then identify biological questions the method can best address in its current or a modified form. Ultimately, no matter how the project is initiated, the overarching motivation for us has been the biological question.

The biological question with which we started our project was: how do individual cells regulate their progression through the cell cycle? However, to collect sufficient data, we needed to address this question using measurements taken from populations of synchronized cells. The "A Primer on Budding Yeast Cell Division and Experimental Methods" sidebar provides additional background details.

In approaching this problem, we needed to learn more about the experimental effects inherent in measuring populations of cells. First, the synchronization procedures used to prepare a population of cells for cell-cycle analysis are not perfect. At best, cells in the population are concentrated near the same approximate point in the cell cycle, but they remain somewhat distributed around this point. Second, individual cells undergo cell division at different rates. So, some time after the beginning of an experiment, one proportion of cells might show markers indicative of a certain level of cell-cycle progression, while other cells might not show these markers at all.

Further complicating the picture, budding yeast cells divide asymmetrically; more specifically, daughter cells take more time to complete G_1 than mother cells. To infer the cell-cycle progression of individual cells, we needed to create a model that not only captured the basic process of cellular reproduction, but also took into account the different sources of synchrony loss that influence measurements from populations of cells. Excitingly, appreciating these aspects of the experimental data collection protocol led to a whole new stream of computational modeling research that we had not imagined at the outset of the collaboration.

As Figure 1 shows, in our CLOCCS (characterizing loss of cell cycle synchrony) model, we represented a population of cells undergoing cell division as a branching process. Branching processes have been widely used in computer science and statistics to model diverse phenomena, from the flow of information in the World Wide Web[3] to the propagation of worms and computer viruses.[4]

In our model, a synchronized cell population at the start of the time-course experiment is positioned at the beginning of a single branch. After completing cell division, this group of progenitor cells gives rise to a new subpopulation of daughter cells, creating a new branch. Over time, as more divisions occur, additional subpopulations join the overall population, starting their own branches of the branching process. At each time point in the experiment, we can estimate the proportion of cells at a particular position in the cell cycle based on the movements of these subpopulations along the branches of the process.

In our formulation of the CLOCCS model, we were mindful of representing cell division with several directly interpretable and informative parameters. Our model learned these parameters from observations of what proportion of cells had visible buds at each time point. These parameters

A PRIMER ON BUDDING YEAST CELL DIVISION AND EXPERIMENTAL METHODS

The cell division cycle is the process by which a cell reproduces itself, replicating its genome and other cellular contents to produce a new cell.[1] The cell division process is traditionally divided into four phases—G_1, S, G_2, and M—but, as Figure A shows, in budding yeast, the latter two are typically merged into a single phase, G_2/M.

The budding yeast is so named because of the way in which this unicellular model organism reproduces. At the transition from G_1 to S phase, the bud—a precursor of the eventual offspring, or daughter, cell—emerges from the progenitor, or mother, cell. During the remainder of the cell division cycle, the bud grows in size. Once the replicated genome and other material are partitioned between the mother cell and bud, the two cells separate and can undergo their own subsequent cell divisions.

To monitor the cell division cycle, experimentalists often use populations of budding yeast cells that have been synchronized so that the cells are temporarily prevented from dividing, paused at (roughly) the same point in the cycle. A variety of synchronization methods exist. One such method involves treating the population with alpha factor, a chemical treatment that causes cells to pause or "arrest" just before the end of G_1. Another common method of synchronization, centrifugal elutriation, preferentially selects small, unbudded cells from a population of cells with the idea that they will be enriched at the earliest points in the cell cycle.

After synchronization, the population is allowed to enter the cell division cycle. Samples of the population are collected at regular intervals over time, during the course of about two cell cycles. Researchers assess each of the samples for the presence or absence of certain cellular features. In the assessment step, the researchers view a small portion of the sample under a microscope or analyze it with other experimental means.

They then record the proportion of cells in the sample that have a particular feature. For example, if a total of 200 cells are examined at a given time, and 40 of them have buds, they would record a budded proportion of 20 percent. The features are specifically chosen because they indicate a cell's position in the cell cycle. For example, the absence of a bud indicates that a cell is in G_1, so 20 percent budded cells means that around 80 percent are in G_1. Features like this that reveal information about a cell's position in the cell cycle are called markers of cell-cycle progression.

Cell-cycle biologists are sometimes interested in how the levels of transcribed mRNAs change over the course of cell division. To assess the (average) transcript level for nearly every gene in the budding yeast genome, biologists isolate RNA molecules from a large population of cells. They process the RNA and then apply it to a DNA microarray. A microarray is a device—usually a glass or silicon chip—on which are arranged a grid of "probes," nucleotide sequences that correspond to individual genes and interrogate the presence of RNA molecules transcribed from that gene.

When the processed RNA is applied to the microarray, each RNA molecule will bind or hybridize to the probe sequence corresponding to the gene from which it was transcribed. In general, the more copies of the gene's transcript, the more hybridization occurs with the gene's probes. Hybridization is usually quantified by a laser, with more hybridization leading to higher fluorescence intensity.

Reference

1. D.O. Morgan, *The Cell Cycle: Principles of Control*, New Science Press, 2006.

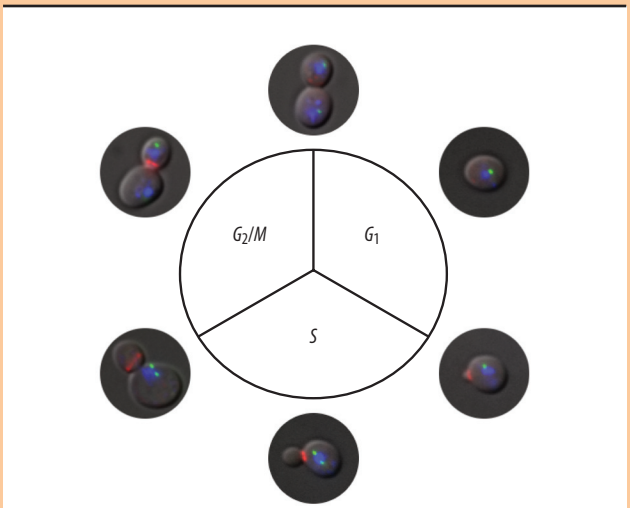

Figure A. The budding yeast cell cycle. The cycle is broken into three main phases (G_1, S, and G_2/M). Along the outside of the pie diagram, overlaid microscope images depict different markers of cell-cycle progression and how they change over the course of cell division. Green: spindle pole bodies; red: myosin rings; blue: DNA.

included recovery time, or the average amount of time each cell spends between synchronization and entry into the cell cycle; cell-cycle duration, or the average amount of time a mother cell takes to complete cell division; the additional time daughter cells take to complete G_1; and the precise time in the cell cycle at which the bud appears. Each of these parameters had a direct biological interpretation and was important for detecting similarities and differences in cell-cycle progression between different strains of yeast or under different environmental conditions.

VALIDATING THE MODEL AND THE BIOLOGY

At each stage of the project, when results or predictions were generated, we took advantage of known biology and the intuition of our biological collaborators to check that our model was sensible. Because our parameters had direct interpretations, we could verify that the parameter values we were learning corresponded to values observed in other experiments. Indeed, our cell-cycle duration inferences were similar to literature-based estimates, as well as to cell division times empirically observed by our collaborators.

In another sanity check, we analyzed data collected from cells synchronized by alpha factor treatment versus centrifugal elutriation. Elutriation tends to put cells under a significant amount of physical stress and is known to extend recovery time. Reassuringly, we found that CLOCCS estimates for recovery time were longer for the elutriated

cells than for cells treated by alpha factor, without significant changes to estimates of cell-cycle duration.

We even went so far as to test the accuracy of our formulation of cell division by simulating DNA content data from our model. In haploid budding yeast, a naturally occurring form preferred by experimentalists for its genetic simplicity, genomic DNA exists in a single copy during G_1, in two copies during G_2/M, and at intermediate levels during S phase. For the purposes of our simulation test, and on the basis of literature evidence, we assumed a linear rate of DNA replication in S phase. Our simulations of DNA content data corresponded well to experimental observations, suggesting that the model was accurately capturing the population's cell-cycle dynamics.

Because our biological collaborators were included in the methods development and verification steps, they proposed many of our most convincing validation tests. These tests were vital to establishing the quality of our models and, thereby, the project's overall success.

Because our biological collaborators were included in the methods development and verification steps, they proposed many of our most convincing validation tests.

IT ALWAYS TAKES LONGER THAN YOU THINK

As we modeled in CLOCCS, synchronized cell populations in a time course are mixtures of cells at different points in the cell cycle, and these mixing proportions vary over time. So, time-series population-level measurements represent average measurements over varying cell mixtures.

Motivating the next phase of our project, our biological collaborators wanted to know how genes were expressed over time at the single-cell level, hoping to gain insight about the molecular underpinnings of cell division. Since budding yeast divide asymmetrically, we were also interested in the different gene expression programs of mother and daughter cells during G_1.

To address these questions, we acquired population-level gene expression measurements using microarrays, and framed the biological questions about single-cell gene expression as a suitable computational task. We recognized that we could use CLOCCS to infer the proportions of cells at different cell-cycle stages at each point in a time-course experiment, and then "de-average" or deconvolve our population-level measurements.

To better illustrate the convolution problem, consider that a single cell's bud is either present or absent over the course of the cell cycle. In a perfectly synchronized population, a plot of the proportion of cells with this binary marker would resemble a square wave, with the transition from all cells unbudded to all cells budded coinciding with the transition from G_1 to S phase. However, due to loss of synchrony in the population and the mixing of cells at different stages of the cell cycle, the proportion of the population with a bud over time instead resembles a damped sinusoid. Likewise, gene expression is a measurement taken from populations of cells and is therefore also subject to these convolution effects.

We developed a model-based deconvolution algorithm to account for these convolution effects and tease apart the gene-expression profiles of mother and daughter cells.[5] The algorithm uses parameter estimates from CLOCCS to determine the proportion of cells in different cell-cycle positions at each point in the time course. These proportions over time are used to form a convolution matrix (H) that maps an average single-cell expression profile (f) to a cell-cycle-averaged population-level expression profile (g).

Because we measured the convolution matrix and population-level expression profile experimentally, the task of deconvolution was to learn the single-cell profile f for each gene. We accomplished this task via regularized optimization, computing the profile f to maximize both the closeness (on a log scale) of the transformed single-cell profile (Hf) to the measured population-level profile g, as well as the smoothness and simplicity of the profile f (using sparse wavelet representations). Our algorithm allowed us to denoise the experimentally observed expression profiles and dramatically increase their dynamic range and temporal resolution, giving us a more precise view of gene expression over time, as shown in Figure 2.

Work on the deconvolution algorithm reinforced some of the lessons we had learned from the first stages of the project and also taught us some new ones. In developing the algorithm, we were dependent on open lines of communication. We were meeting just as regularly as before, checking the plausibility of any assumptions of the algorithm with our biological collaborators and discussing data idiosyncrasies. For example, since gene-expression measurements from microarrays often involve multiplicative (rather than additive) Gaussian error, our constrained optimization had to minimize the squared distance between the transformed single-cell profile (Hf) and the population-level profile g on a logarithmic scale.

In addition to maintaining our commitment to frequent communication, we also learned to appreciate the time required to do principled work. Biological experiments are sometimes finicky—for example, isolation of highly unstable RNA for microarrays—and more complex experimental procedures do not always work without tedious trial and error. The more we conferred with our biological counterparts, the more respect we developed for the practical aspects of the experiments.

Likewise, our biological collaborators might have been surprised at how long it takes to get modeling details just

right, were it not for their frequent interactions with the computational members of the project.

ART AND DIPLOMACY IN MODELING

As all modelers will attest, modeling often requires making simplifying assumptions about the problem domain. This has proven no less true in our collaboration. However, we now have an added consideration when we decide which assumptions to make: what do our collaborators think of the assumptions? We are constantly striking a balance between computational and mathematical convenience on one hand, and biological reality on the other. We always include our biological collaborators when we make modeling choices. We confer when deciding the level of abstraction in our modeling assumptions, and set the level largely by determining the point beyond which the biological plausibility or interpretability of the model begins to suffer.

Striking this balance between computational convenience and biological reality in the modeling arose at every stage of our project, particularly when we started to consider markers of cell-cycle progression beyond just the presence of a bud. We were still concerned with the overarching biological question of how cells regulate their progression through cell division but wanted to attain a higher-resolution view of the process.

We noted that different markers carry different information about a cell's progress in cell division. One such marker is the amount of DNA each cell contains. We therefore extended our model to use DNA content information: a cell is in G_1 if it has one genome copy, is in G_2/M if it has two genome copies, and is proportionally through S phase if somewhere in between. Using DNA content in our model helped us "chop" the cell cycle into more pieces and increase our resolution of a cell's position in cell division.[6]

We also had available to us still other markers that distinguish cells in finer intervals of the cell cycle. For example, we fused fluorescent proteins to proteins known to change their status during cell division.[7] In brief, whenever the cell-cycle-regulated protein is expressed, it can be seen because of the fluorescent protein attached to it. With the aid of a fluorescence microscope, we could then visualize the proportion of cells with a particular marker status at each point in time.[8] One of our fluorescent markers is the spindle pole body (SPB), a structure that is important for partitioning the duplicated genome between the mother cell and the newborn daughter. Another fluorescent marker is the myosin ring, which appears on the periphery of the mother cell late in G_1, indicating the site where the bud will soon emerge. The myosin ring disappears with the completion of cytokinesis.

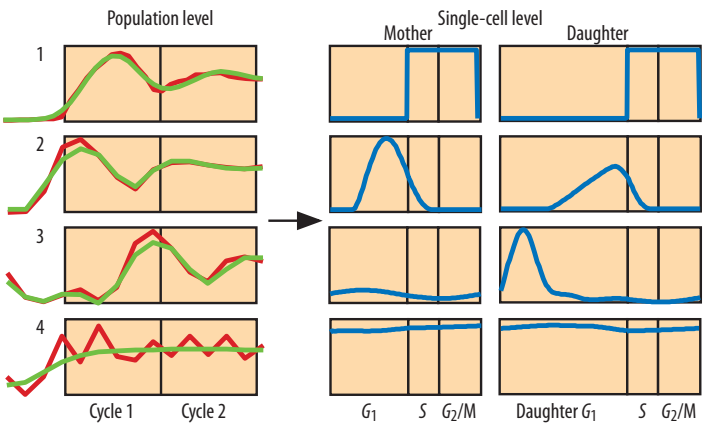

Figure 2. Deconvolution and its effect on measurements in cell populations. The diagram shows different types of measurements for budding yeast at the population level and the effect of deconvolution on distinguishing single-cell patterns from these measurements. The figure highlights some of the benefits of the deconvolution algorithm. (1) Budding measurements. (2) Gene expression patterns shared by mother and daughter cells. (3) Gene expression patterns specific to daughter cells. (4) Gene expression patterns denoised by deconvolution.

In assessing cells under a fluorescence microscope, we counted the myosin ring as a binary feature of the cell—that is, cells either had myosin rings or not. The SPB exists in a single copy for most of G_1. During late G_1, however, the SPB duplicates. From late G_1 to the middle of G_2/M, the two SPBs separate from one another, but remain within the mother cell—a "short spindle." Finally, from the middle of G_2/M to the end of cytokinesis, the two SPBs separate further, one remaining in the mother cell and the other moving into the newborn daughter cell—a "long spindle."

At each time point, we examined samples of the population under the microscope and counted the proportion of cells with buds, myosin rings, short spindles, and long spindles. Because we knew that available markers could vary from experiment to experiment (and from lab to lab), we generalized CLOCCS to incorporate any arbitrary number or combination of binary markers. To do this, we introduced new parameters indicating the points in the cell cycle at which each of the different binary markers appeared and disappeared.[9] As we had done previously with bud counts, we could estimate our marker-specific parameters from the corresponding marker observations.

We could have modeled these markers differently. For example, rather than considering a cell as either having a bud or not, we could have used the microscope images to generate continuous measurements of bud size. Likewise, for some of the other fluorescence-based markers like the long and short spindles, we could have measured the approximate distance between the SPBs instead. Modeling

these markers as continuous-valued cellular features is certainly a possible future direction for our project. We chose a coarser level of abstraction as a starting point for the model and as a foundation for further development.

The binary representation of the microscope-derived markers also led to biologically interpretable parameters related to the points in the cell cycle at which the markers appeared and disappeared. Regardless of how the markers were modeled, what was important for us was that the decision was made with the input of both computational and biological members of the project.

With the rapid generation of new datasets and biological insights, systems biology has great potential for broader understanding of biological processes. Numerous opportunities exist for development of computational models and analysis tools to extract biological insights from these data.

For us, making the most of these opportunities has depended upon our commitment to open and frequent communication between computationalists and biologists. In initiating our collaboration, we focused on the biologists' goals, using biological questions as the drivers of downstream method development. We used these frequent initial interactions to help break down cultural barriers, establish common goals for the research, and give each side a greater appreciation for the details of the other side's work. The more informed we became about each other's perspectives, the more adept we became at proposing new directions for the project.

We checked the model's potential assumptions with one another. We turned to each other when our model developments were drifting away from biological reality and when we wanted to develop more sophisticated models of the experimental effects and biological phenomena. Through all the frustration, we continued to meet face to face, and ask questions of one another. We continued to find new and better ways of describing our experimental techniques and model developments.

Indeed, the past few years have shown us that biological and computational researchers can share common goals when they dedicate themselves to finding a common language for frequent, effective communication.

Acknowledgments

This work was funded by grants from the NIH (P50-GM081883-01 to S.B.H. and A.J.H., inter alia) and from DARPA (HR0011-09-1-0040 to A.J.H.).

References

1. T.R. Cech and G.M. Rubin, "Nurturing Interdisciplinary Research," *Nature Structural & Molecular Biology*, Dec. 2004, pp. 1166-1169.
2. H. Kitano, "Systems Biology: A Brief Overview," *Science*, Mar. 2002, pp. 1662-1664.
3. D. Wang et al., "Information Spreading in Context," *Proc. 20th Int'l Conf. World Wide Web* (WWW 11), ACM, 2011, pp. 735-744.
4. S.B. Sellke, N.B. Shroff, and S. Bagchi, "Modeling and Automated Containment of Worms," *IEEE Trans. Dependable and Secure Computing*, Apr. 2008, pp. 71-86.
5. X. Guo et al., *Branching Process Deconvolution Algorithm Reveals a Detailed Cell-Cycle Transcription Program*, submitted for publication, 2012.
6. D.A. Orlando et al., "A Branching Process Model for Flow Cytometry and Budding Index Measurements in Cell Synchrony Experiments," *Ann. Applied Statistics*, Dec. 2009, pp. 1521-1541.
7. M.S. Longtine et al., "Additional Modules for Versatile and Economical PCR-Based Gene Deletion and Modification in Saccharomyces cerevisiae," *Yeast*, July 1998, pp. 953-961.
8. J.W. Lichtman and J.-A. Conchello, "Fluorescence Microscopy," *Nature Methods*, Dec. 2005, pp. 910-919.
9. M.B. Mayhew et al., "A Generalized Model for Multi-Marker Analysis of Cell Cycle Progression in Synchrony Experiments," *Bioinformatics*, July 2011, pp. 295-303.

Michael B. Mayhew *is a PhD candidate in the computational biology and bioinformatics program at Duke University. His research interests include Bayesian statistical modeling, image processing, and time-series analysis. Mayhew received an MSc in computer science from McGill University, Canada. Contact him at michael.mayhew@duke.edu.*

Xin Guo *is a PhD candidate in computer science at Duke University. His research interests include computational biology and statistical learning. Guo received an MSc in informatics from Saarland University, Germany. Contact him at xin.guo@duke.edu.*

Steven B. Haase*, an associate professor of biology at Duke University, found evidence for a CDK-independent cell-cycle oscillator. His research interests focus on the mechanisms regulating the temporal events of the cell division cycle. Haase received a PhD in genetics from Stanford University. Contact him at shaase@duke.edu.*

Alexander J. Hartemink*, a distinguished associate professor of computer science, statistical science, and biology at Duke University, uses Bayesian statistics and machine learning to address problems in molecular systems biology and genomics. His research focuses on understanding the mechanisms of transcriptional regulation and its role in controlling the cell division cycle. Hartemink received a PhD in electrical engineering and computer science from MIT. Contact him at amink@cs.duke.edu.*