acmqueue Metamorphosis: the Coming Transformation of Translational Systems Biology

In the future computers will mine patient data to deliver faster, cheaper healthcare, but how will we design them to give informative causal explanations? Ideas from philosophy, model checking, and statistical testing can pave the way for the needed translational systems biology.

Samantha Kleinberg and Bud Mishra, New York University

One morning, as Gregorina Samsa was waking up from anxious dreams, she discovered that she had become afflicted with certain mysterious flu-like symptoms that appeared without any warning. Equally irritating, this capricious metamorphosis seemed impervious to a rational explanation in terms of causes and effects. "What's happened to me?" she thought. Before seeing a doctor, she decided to find out more about what might ail her. She logged on to a Web site where she annotated a timeline with what she could remember. Since March, she'd had more headaches than usual, and then in April she had begun to experience more fatigue after exercise, and as of July she had also experienced occasional lapses in memory. "Why don't I go back to sleep for a little while longer and forget all this foolishness," she thought. As she was about to abandon this errand, the system came back to life with a barrage of questions: Is she female? Had she experienced any significant stress in the past few months? Had she noticed any joint or muscle pain? It also obtained her permission to download her genomic profile.

After reviewing Mrs. Samsa's history and the other information gathered, the system suggested she might have chronic fatigue and that she suffered from a particular category of chronic fatigue. Given her similarity to other patients, the system showed the trajectory her illness had taken, and that it likely began in early February. Since other patients' histories and genomic profiles—including their recoveries—were known and had already been analyzed, it predicted what lay ahead for her, as well as her best course of treatment. Mrs. Samsa (unlike her namesake from the last century) was lucky; it was the year 2019, and she had access to personalized medicine that grew out of many years of genomics technology development, as well as computational and systems biology research. (The accompanying sidebar lists current Web sites about personalized medicine.)

Now imagine you are a present-day researcher engaged in patient studies to understand CFIDS (Chronic Fatigue and Immune Dysfunction Syndrome). Most likely, your data come from doctors who observe the course of the illness in a few patients, but who mostly gather information, noting events these patients describe. These histories are incomplete and subject not only to missing but also inaccurate information. A patient may not be sure when his or her fatigue started, whether it had a sudden onset or a slow progression, or if it might have been accompanied by a diagnosis of depression. These doctors' own theories about what causes the disease influence the medical tests they recommend. From these partial medical histories and small sample of patients, you wish to understand the disease etiology, a word with a Greek root (*aitia*) meaning cause.

In medicine, etiology now stands for the science that deals with the causes or origin of disease, the factors that produce, or predispose one toward, a certain disease or disorder. The etiology of

CFIDS has been recalcitrant, heterogeneous, and complex: it is now believed that the underlying causal relationships may not be the same for all patients, and that there may be different forms of the syndrome that have different causes and progress according to different rules. The ability to categorize a patient and understand the finer details of his or her particular condition is needed for personalized medicine, where therapeutic interventions are tailored to each patient.

This general problem is somewhat severe for CFIDS but is not unique to it. We could have just as easily described a population in sub-Saharan Africa where more than a million people, most of them children, die each year from a deadly drug-resistant form of malaria transmitted by the parasite *Plasmodium falciparum*. Traditionally, biologists would focus on precisely describing one small piece of the whole organism—perhaps just one metabolic pathway or the interactions of a few proteins. The approach of systems biology is to understand the workings of the entire system as a whole and view the individual parts within the context of that whole. Such knowledge would be extremely useful in targeting specific genes and proteins in pathways common to many different strains of *P. falciparum* in order to design effective vaccines.

Within systems biology, there are two competing (possibly, ultimately complementary) approaches aimed at disentangling the myriad of causal relationships. Each approach operates locally, but together they affect the whole system to determine its global properties. These two approaches can be classified as bottom-up (mechanistic/molecular) or top-down (phenomenological/causal). The mechanistic/molecular approach creates models that allow detailed simulation of various changes and events; these models are judged by how well they predict a system's response to the changes. In this case one begins with a set of mathematical models of how things work and then tests the predictions of the models experimentally. In contrast, the phenomenological/causal approach starts with known theories and observations, and generates new refutable hypotheses, driving the search for agents and mechanisms through new experiments.

The mechanistic/molecular approach is deductive, and its soundness depends on its prior beliefs about the basic mechanisms, as well as the completeness of the priors. The phenomenological/causal approach is deductive, but its completeness depends on the dimension, size, and resolution (temporal and spatial) of its observations. In what follows, we describe the new mathematical and algorithmic tools that we have been developing for phenomenological and causal analysis of biological systems: ecologies, populations, organisms, cells, pathways, and molecules.

Personalized Medicine on the Web

Disease social networks:	www.patientslikeme.com
Genetic testing:	www.23andme.com
	www.decodeme.com
	www.navigenics.com
Whole genome sequencing:	www.knome.com
Medical history and genetic info:	www.personalgenomes.org
Medical search engine:	www.medstory.com

CAUSAL RELATIONSHIPS

UNDERSTANDING CAUSALITY

Let us begin with the following fundamental questions: What is a causal relationship? What distinguishes causation from mere correlation? The majority of work on causality is in philosophy, where philosophers have distilled out just a handful of fundamental ways to describe causal relationships, one of them being in terms of probabilities. (Others include *counterfactuals*[17] and *processes*.[4,22]) In the "probabilistic theories of causality," a cause simply occurs earlier than the effect and raises the probability of the effect, and a causal relationship between *a* and *b* means that either *a* causes *b* or *b* causes *a*.[26]

A number of events or factors, however, may falsely appear to be causes under this simple definition. For example, a barometer reading normally falls before it rains, and the probability of rain given that the barometer reading has fallen is greater than just the probability that it will rain. Saying that a barometer read-out causes rain, however, would be rather foolish. This problem is resolved by taking into account their common cause: decreasing air pressure is the actual cause of both the barometer reading falling and the rain. To identify and avoid such erroneous (or spurious) causes, the relationship between cause and effect must be more precisely defined.

A spurious cause is one that may appear indistinguishable from genuine causes—by seemingly raising the probability of the effect—but in fact has little or no impact on the occurrence of the effect. One way of recognizing such a cause is to look for other, earlier causes that account for the effect at least as well as the potentially spurious cause does. In the case of the barometer, the air pressure drops before the barometer falls, and once we know that the air pressure has dropped, the falling barometer does not add any information. We do not necessarily wish to impose such a strict condition—perhaps only that the barometer adds little information (as measured by the difference in conditional probabilities with and without this event). Further, we do not necessarily wish to look only for earlier conditions or be satisfied with calling a cause spurious only if we can detect one other cause that better explains the effect. Instead, we may wish to estimate the difference a cause makes with respect to all other putative causes of the effect, by averaging the difference in probabilities, and use this to distinguish spurious causes from genuine ones. (For more information on the philosophical issues with detecting spuriousness, see Suppes[26] and Eells.[6]) Because of these subtleties, representation, definition, and the algorithmic search for causality must be carried out with precision.

REPRESENTING CAUSALITY

In computer science, a primary method for causal inference has been through the use of graphical models such as Bayesian networks and dynamic Bayesian networks, which allow for systems with a temporal component. (For more information on graphical models in causality see Pearl[21] and Spirtes et al.[23] Detailed descriptions of dynamic Bayesian networks may be found in Friedman et al.,[10] Murphy and Mian,[19] and Langmead.[16]) In all cases, the model is defined by a graph where edges between nodes represent dependencies and missing edges represent independence between nodes. The models represent the joint probability distributions of the system and can be inferred by testing for conditional independence. (Finding causal models is somewhat more complex, particularly in the case of data without a temporal component. In that case, the edges of the graph are oriented based on the conditional independence relations found and assumptions about the system.) In these

models, there is no natural way to reason about or test for complex causal relationships, and we do not want to restrict our methodologies to deal with just the cases of one event causing another event.

Consider the warning we frequently hear: "Smoking causes lung cancer." Cigarette package warnings in the UK go even further with statements such as "Smoking kills" and "Smoking causes fatal lung cancer." How would you interpret these statements? They implicitly deal with two concepts—namely, the probability of getting lung cancer if you smoke and how long it will take for you to get lung cancer. If the probability is nearly one, but it will take many decades, you may make very different decisions about smoking than if the probability is perhaps a quarter, but lung cancer develops within 10 years. For this reason we must be able to infer and represent the time over which a cause produces its effect, as well as the probability that it will do so. A convenient and efficiently computable formalism for a probabilistic definition of causality (which also incorporates temporal priority) is already available in *probabilistic temporal logic* and the efficient *model-checking* algorithms that have been developed for this logic.

Temporal logic allows us to reason about not just *if* a formula is true, but *when* it is true. That is, instead of a proposition being only *possibly* or *necessarily* true, we can make statements such as, "If the barometer falls, at the next state it will rain." Probabilistic temporal logic, such as PCTL (probabilistic computational tree logic),[11] extends this idea to include statements with probabilities, as well as explicit deadlines. We could then state, "If the barometer falls, then it will rain within one to two days with probability 0.9." The truth values of formulas in PCTL are defined relative to probabilistic structures. These consist of a set of states, labels describing the properties true within the states, and transition probabilities that tell us which states can be followed by other states and with what probabilities. The state labels are from a set of atomic propositions—such as events or facts—that will be the building blocks of causes and effects.

More formally, we have two types of formulas: *state* and *path*, just as in classical CTL. A state formula, such as $p \wedge q$, is true in a particular state if the properties hold in the state (i.e., if it is labeled with both p and q). A path formula refers to a sequence of states along which a formula must hold. The set of valid formulas can be defined by the following:

1. Each atomic proposition is a state formula.
2. If f_1 and f_2 are state formulas, so are $\neg f_1$, $(f_1 \wedge f_2)$, $(f_1 \vee f_2)$, $(f_1 \rightarrow f_2)$.
3. If f_1 and f_2 are state formulas, and t is a nonnegative integer or ∞, then $f_1 U^{\leq t} f_2$ and $f_1 \mathcal{U}^{\leq t} f_2$ are path formulas.
4. If f is a path formula and p is a real number with $0 \leq p \leq 1$, then $[f]_{\geq p}$ and $[f]_{>p}$ are state formulas.

The "Until" formulas in item 3 are path formulas, as they refer to sequences of states. The first (involving *strong until*), $f_1 U^{\leq t} f_2$, means that f_1 will hold at every state *until* f_2 holds, which will happen in less than or equal to t time units (where each transition is considered to take one unit of time). The second formula (involving *weak until*) is similar, except with the condition that f_2 is not guaranteed to hold. When that is the case, f_1 must hold for a minimum of t time units. Item 4 adds probabilities to these path formulas, turning them into state formulas. For example, if we take:

$$[f_1 U^{\leq t} f_2]_{\geq p}$$

(which may be abbreviated as $f_1 U^{\leq t}_{\geq p} f_2$), this means that with probability at least p, f_2 will become true within t time units, and f_1 will hold along the path until that happens. This state formula is evaluated by calculating the probability over the set of possible paths from the state, where the

4

probability of a path is the product of the transition probabilities along the path. We will also use an operator called "leads-to," written:

$$f_1 \rightsquigarrow_{\geq p}^{\geq t_1, \leq t_2} f_2 \equiv AG\left(f_1 \rightarrow F_{\geq p}^{\geq t_1, \leq t_2} f_2\right)$$

which is interpreted as: from a state where f_1 holds, f_2 will hold in between t_1 and t_2 time units, with probability p. We have added a lower time bound to the standard leads-to operator, as described in Hansson and Jonsson,[11] in order to accommodate the temporal priority condition for causality.

DEFINING CAUSALITY

Causal relationships can now be defined in terms of PCTL formulas. (*Defining* is something of a misnomer here. This is what we will be referring to when we say causal and what our algorithms will identify; however, these do not define what it is for something to be a cause or suggest that true causes must fit all of these criteria. There will be causes we do not identify by this method, but we do not say that these are not causes; rather, that they must be inferred by other means.)

A reader uninterested in the formal treatment of these topics, however, may skip the mathematical sections and go directly to an overview of the procedure, found at the end of the following section, and then continue on with the applications in the section titled "A Short Example." First, possible (also called *prima facie*) causes are those that are earlier than and raise the probability of their effects. This means that c is a possible cause of e if c and e are valid PCTL formulas and all of the following are true:

1. $F_{>0}^{\leq \infty} c$

2. $c \rightsquigarrow_{\geq p}^{\geq 1, \leq \infty} e$

3. $F_{<p}^{\leq \infty} e$

The first item simply says that c occurs with nonzero probability. The second means that after the possible cause, the effect occurs with at least probability p (we can also further restrict this time window when background knowledge makes it possible, but the minimum condition is that c is earlier than e by at least one time unit). The third says that the probability of the effect is less than p; thus, c raises the probability of e.

As noted before, this definition admits many erroneous causes. To assess the significance of a particular c as a cause for a particular e, we compute the difference that c makes for e with respect to all other possible causes of e. These other causes may be at any time prior to e (they could be at the same time as or even after c). We begin with X the set of prima facie causes of e. Then, for each $x \in X \backslash c$, we compute:

$$\varepsilon_x(c,e) = P(e|c \wedge x) - P(e|\neg c \wedge x)$$

The average of these gives:

$$\varepsilon_{avg}(c,e) = \frac{\sum\limits_{x \in X \backslash c} \varepsilon_x(c,e)}{|X \backslash c|}$$

For each of the possible causes, we have now computed its average potency as a predictor of its

effect. We then use this ε_{avg} to determine c's significance, where causes with values less than some threshold are considered insignificant. We will discuss how to choose this ε in a rigorous way, but first we provide an overview of the process of going from data to causal hypotheses.

INFERRING CAUSALITY

How would the inference process work? Imagine trying to learn how to use a microwave oven by observing how other people use one. The oven can have its door open or closed, its switch turned on or off, and may or may not have food placed in it. Most of the time the oven is in a "dormant state," with its door closed, its switch off, and containing no food; but the oven can also enter an "active state" when the door is closed, the switch is on, and it contains food. There are impossible states (e.g., the door is open with the switch on), as well as possible but futile states (e.g., the door is closed, the switch is on, but no food is in the oven). From the dormant state, the oven is most frequently observed moving into the active state by having its door opened, food placed inside, the door closed, and finally the switch turned on; and it reverts to a dormant state by having its switch turned off, door opened, food removed, and door closed. We may observe, on a few rare occasions, the oven operated in an unusual manner. Each of our observations shows us a possible state of the system (a collection of properties that are true at a particular time).

If the underlying structure is of the probabilistic kind described earlier, what we have observed is one possible path through the system being studied. We have seen a sequence of states that the system is allowed to traverse (with some probability), and because of the time order we can infer which states may follow others. From the frequency of these observations, we can estimate the transition probabilities. Thus, from the data we can determine which of the properties of interest have been satisfied when subject to model checking (see Clarke et al.[3]). Of the properties that are satisfied, we then know which are prima facie causes based on their probabilities and the probabilities of their effects, and we can compute for each. Note that this is a computationally intensive task. In the worst case, when every item is a possible cause of every other item, computing these epsilons is $O(N^3T)$, where N is the number of items and T the number of time points. The worst-case space complexity is comparatively much smaller, $O(NT + N^2)$, though still large relative to the dataset.

Once we have tested our causal hypotheses and computed the strength of each, we can subject them to statistical tests to determine which are significant. If we were testing a single hypothesis, we could accept or reject it based on the probability of seeing such a result under the null hypothesis (e.g., there is no causal relationship). When we are testing a multitude of hypotheses, however, we increase our chances of seeing anomalous behavior and thus risk making errors in our testing if we do not account for this. To maintain the same significance level (probability of rejecting the null hypothesis when it is true) when conducting multiple tests, we must adjust how we reject the null hypothesis. (For more on hypothesis testing and the false discovery rate, see Storey and Tibshirani,[25] Efron,[7] and Benjamini and Yekutieli.[1]) We focus on controlling these errors, rather than those when we fail to make a discovery (falsely accepting the null hypothesis). Intuitively, when we are testing a small number of hypotheses, making even a few false discoveries is undesirable; when we are testing thousands of hypotheses, however, we can be more tolerant and just try to control the *proportion* of false discoveries.

One method of controlling the FDR (false discovery rate)[7] uses an *empirical null*, freeing us from

having to specify a null hypothesis—instead, it is estimated from the data. This method also controls the FDR in cases with dependency, which is particularly useful in datasets with complex causal structures. Beginning with N hypotheses and their test statistics $z_1, z_2,…,z_n$ (a z-value, also called a standard score, denotes how many standard deviations a value is from the mean), we assume that our tests fall into two classes: "interesting" and "uninteresting," corresponding, respectively, to rejection and acceptance of the null hypothesis. We implicitly assume that the number of interesting hypotheses is small relative to $|N|$. The prior probabilities of a test (here, a causal hypothesis) falling into the uninteresting and interesting classes are respectively p_0 and $p_1 = 1 - p_0$, where p_0 is large (perhaps 0.99). The densities $f_0(z)$ and $f_1(z)$ describe the distribution of these probabilities. When using a theoretical null, $f_0(z)$ is the standard $N(0,1)$ density. Note that we need not know $f_1(z)$. We define the mixture density:

$$f(z)=p_0 f_0(z) + p_1 f_1(z)$$

Then the posterior probability of a case being uninteresting given z is:

$$Pr\{null|z\} = p_0 f_0(z) / f(z)$$

and the local false discovery rate is:

$$fdr(z) \equiv f_0(z) / f(z)$$

Note that p_0 is omitted. We may estimate this value and include it in the calculations, but since we assume it to be nearly one, this does not grossly overestimate the *fdr*.

Thus, the overall procedure is as follows:
1. Enumerate logical formulas, using background knowledge or testing those up to a specified level of complexity.
2. Test which of the formulas are satisfied by the system and satisfy the conditions for prima facie causality.
3. For each prima facie cause, compute the associated ε_{avg}.
4. Translate values for ε into z-values.

Software Challenges
- The most intensive computation (that of the epsilon values) could be parallelized to run simultaneously on many nodes, since each result is independent.
- Visualizing the results is a difficult task in cases involving: many elements; a high degree of dependency; or both. We need to develop graphical representations that capture the complexity, as well as the temporal component of the inferred relationships.

5. Estimate $f(z)$ from the observed z-values.

6. Define the null density $f_0(z)$ either from the data or using the theoretical null.

7. Calculate $fdr(z)$ using the equation $fdr(z) \equiv f_0(z) / f(z)$.

8. For each causal hypothesis where $\varepsilon_{avg}(c,e)$ corresponds to z_i such that $fdr(z_i)$ is less than a threshold (say, 0.01), label it as significant and the rest as insignificant.

   The advantage of this procedure is that we do not need to specify an arbitrary threshold to determine the difference a cause must make to its effect in order to be considered significant. Rather, we can test a multitude of arbitrarily complex causes and use a rigorous statistical method (making minimal assumptions about the data) to determine the significance of each hypothesis. Finally, note that all of these methods can be applied to temporal datasets in a variety of domains (biology, finance, etc.). Our choice of Common Lisp allowed rapid prototyping and implementation, as well as the ability to handle the heterogeneous data.
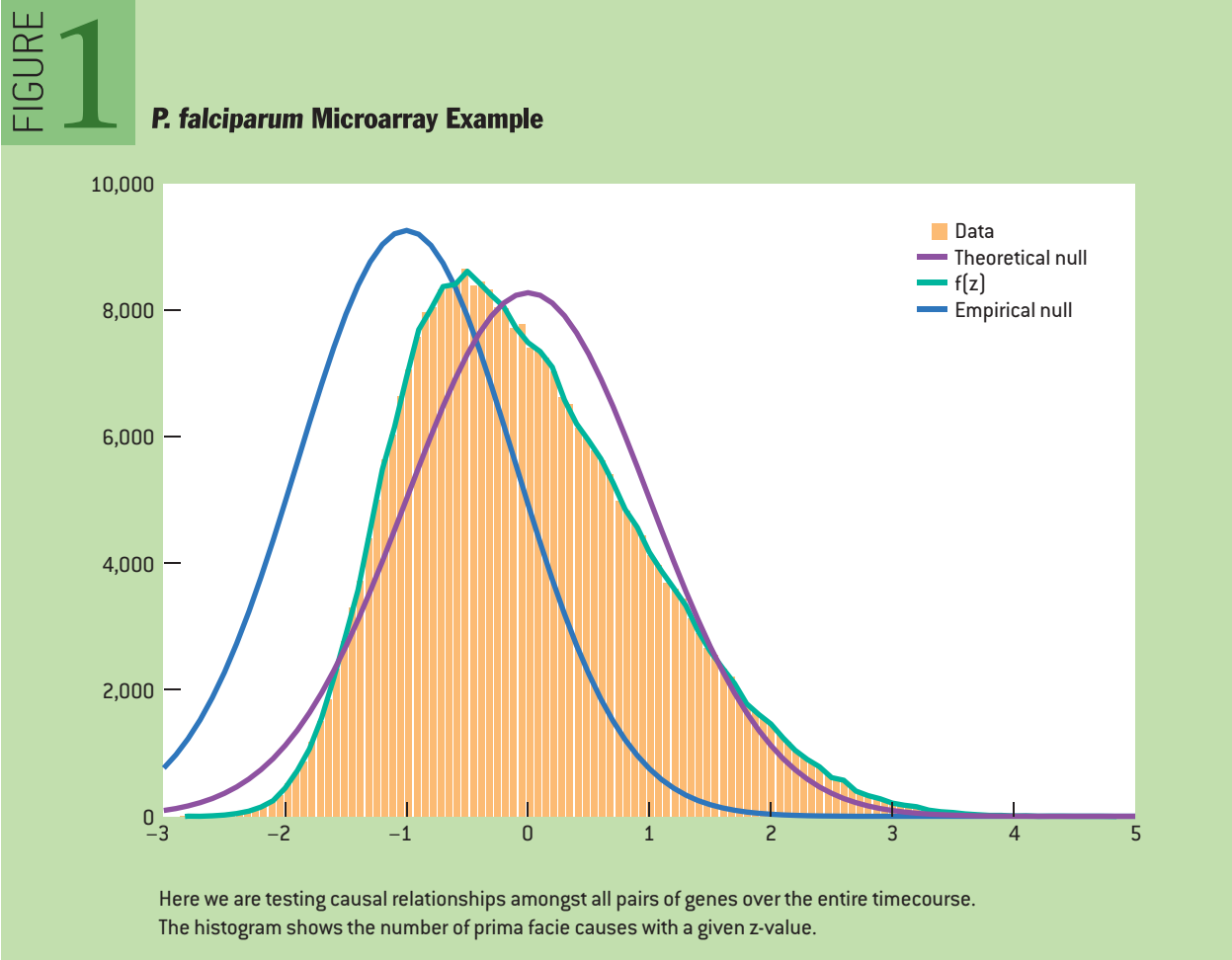
A SHORT EXAMPLE

Let us revisit the case of *P. falciparum*, the parasite responsible for the deadliest form of malaria. Using the system described earlier, we analyzed a set of time-course gene-expression data, collected with microarray technology, covering *P. falciparum*'s 48-hour IDC (Intraerythrocytic Developmental Cycle).[2] Microarray data, where expression levels may be measured for thousands of genes at a time, have been the subject of many studies on both multiple hypothesis testing,[5,8] as well as causal inference.[9,12,19,20,24]

   To narrow the focus a bit, let's take the stage of *P. falciparum*'s life cycle that is responsible for most malaria symptoms and look at how to decipher the rules governing this IDC, as well as how to exploit this information to identify potential targets for drugs or vaccines. Although the *P. falciparum* genome has been sequenced, only about half of its genes have functional annotations. Because of its low sequence similarity with other organisms and even other strains of *Plasmodium*, it is difficult to propose new annotations. In related work we have combined multiple sources of data to predict annotations better, but found no way to guide our method toward genes that might be of particular interest—a common problem shared by many ontology-based bioinformatics tools.[18] By inferring causal relationships, we have a better chance of finding key genes, whose annotations point to novel experimental studies and may better reveal the complex mechanisms underlying this developmental cycle.

   The IDC has three main stages, with all genes active at some point during the cycle, forming a so-called "cascade" of activity. Rather than looking only at relationships across the entire time course, we also used the critical time points we found in our earlier work[15] to segment the data into five windows of non-uniform size. These windows correspond to the three stages plus the periods of transition between the stages. In each window, we generated many putative relationships, using the PCTL formulation, primarily by considering all pairs of genes, where the influence of a gene on another is assumed to last, at most, one unit of time. In other words, we considered all formulas of the form:

$$c \leadsto^{\geq 1, \leq 1} e$$

where c and e represent the under- or over-expression of particular genes. Note that in practice we encounter a massive number of potential hypotheses but only a relatively small number of

FIGURE 1

P. falciparum Microarray Example

Here we are testing causal relationships amongst all pairs of genes over the entire timecourse.
The histogram shows the number of prima facie causes with a given z-value.

time points with which to interpret them, and we expect most of the putative hypotheses to have occurred by happenstance and belong to the null model.

After examining the entire time course, we decided to restrict our dataset to only genes known to be involved in protein-protein interactions, leaving us with $N=2,846$ unique genes. In all cases, we computed the possible causes and their epsilons using our Common Lisp implementation of the previously described methods and computed the empirical null using the method of Jin and Cai with the R code they have made available.[13]

In figure 1, notice that practically all the data falls within the plotted theoretical null $N(0,1)$, and thus (under that null hypothesis) reveals nothing very interesting. The empirical null $N(-1.00, 0.89)$ corrects for the bias, as well as the positive skew (long tail to the right) and is thus shifted much further over than the theoretical null.

Note that we see an overabundance (in thousands) of prima facie causes where $f(z_i) < 0.01$—a fact that warrants some biological explanation. It is believed that biological systems are, by necessity, quite robust,[14] consisting of a large number of correlated cause-and-effect relationships that orchestrate the systems dynamics in a fail-safe manner. In this particular example, we see that during each phase of the IDC, the genes related to that phase act in concert, producing the cascade and creating numerous correlated dependencies: (1) many genes are causally related to many other

genes that are active in the same IDC stage and are organized in a complex network of interactions; (2) there are many backup mechanisms to allow the cascade to continue uninterrupted in case of some perturbation to the system. Thus, we speculate that most biological systems, like this one, exhibit two classes of causal relationships: genuine primary causes and genuine backup causes.

The most interesting windows we found were the second and fourth (hours 7-16 and 28-43), which correspond to transitions between IDC stages and a major reorganization of regulatory activities. The second window spans the end of the first stage to the start of the second stage. Similarly, the fourth window spans the end of the second stage and most of the third stage (during this time it begins, takes place, and starts to wind down). Figures 2 and 3 show histograms of the values of ε_{avg} as computed for all prima facie causes found in those windows. The empirical null for the second time window is given by $N(-1.32, 0.71)$, differing substantially from the theoretical null $N(0,1)$. For the fourth window, the empirical null $N(-0.62, 1.13)$ is similar to the theoretical null, though shifted to the left. In both cases, thousands of prima facie causes would be deemed statistically interesting, or genuine, based on their computed ε_{avg}. That is, they had low fdr values.

To find out how much the temporal dynamics influenced the results, we compared these results with those found in the earlier analysis of the whole time course. We then took the most significant hypotheses from windows 2 and 4 and the entire time course (totaling 5,996, 8,087, and 1,963, respectively) and further analyzed their intersection. The results are shown in table 1. Despite the large number of hypotheses, we see very little intersection between the two time windows. Further studies are needed to detect whether those common to both windows are significant or, more likely, housekeeping genes whose activity is consistent throughout the IDC. We visualize the networks of these hypotheses for the two windows of time, as shown in figures available at: http://bioinformatics.nyu.edu/~skleinberg/supp/dgc.shtml.

Table 1.

Data	Intersection size
Windows 2 and 4	117
Window 2 and entire time course	285
Window 4 and entire time course	347
Windows 2,4, and entire time course	25

CONCLUSION

We are building a software tool that is primarily focused on understanding disease etiologies; interrelations among various diseases; relations of a disease to polymorphisms, biomarkers, and pathways; and, finally, approaches to use this knowledge to select diagnostics, prognostics, and therapeutics for patients in a personalized manner and with high precision. These approaches have their genesis in the methodologies we had developed in GOALIE,[15] which took time-course gene-expression data, and from the data generated relevant qualitative explanations of the biological systems, either formally described in a temporal logic or expressed in natural language. In these ways, these tools could contribute enormously to progress in both basic biological sciences, as well as their applications to biomedical fields.

Two problems need to be better understood. First, when a new patient arrives with a single

FIGURE 2

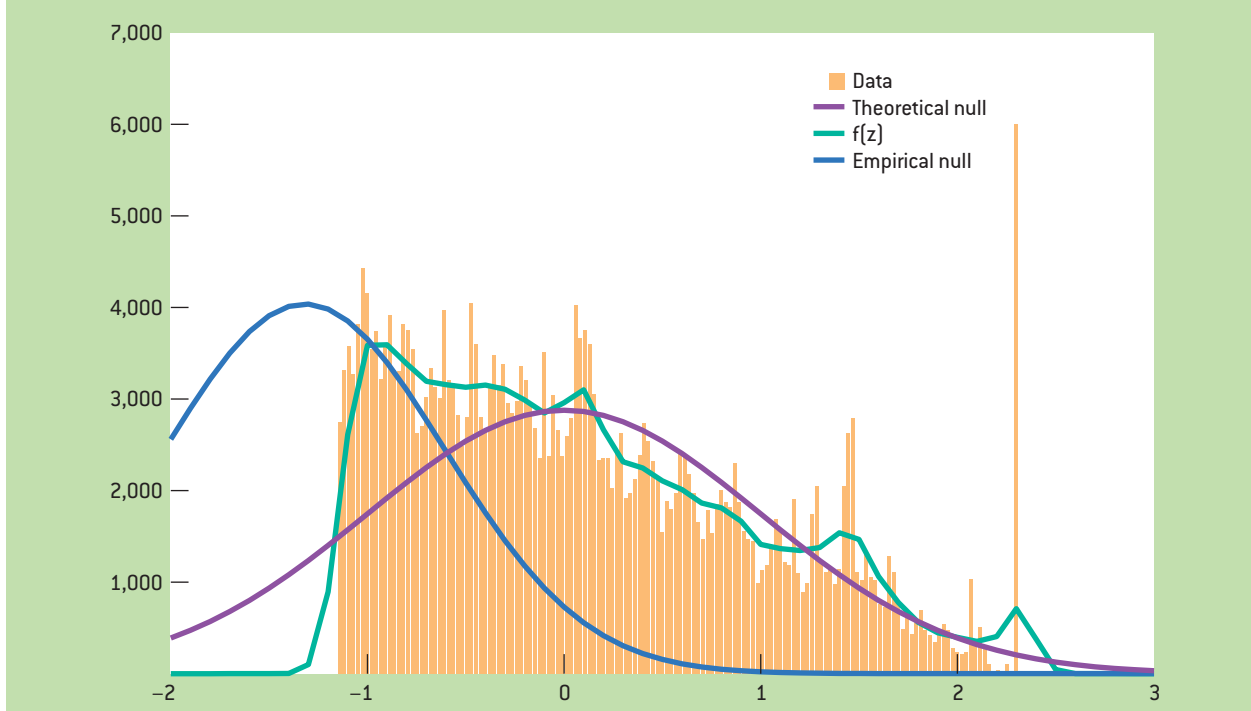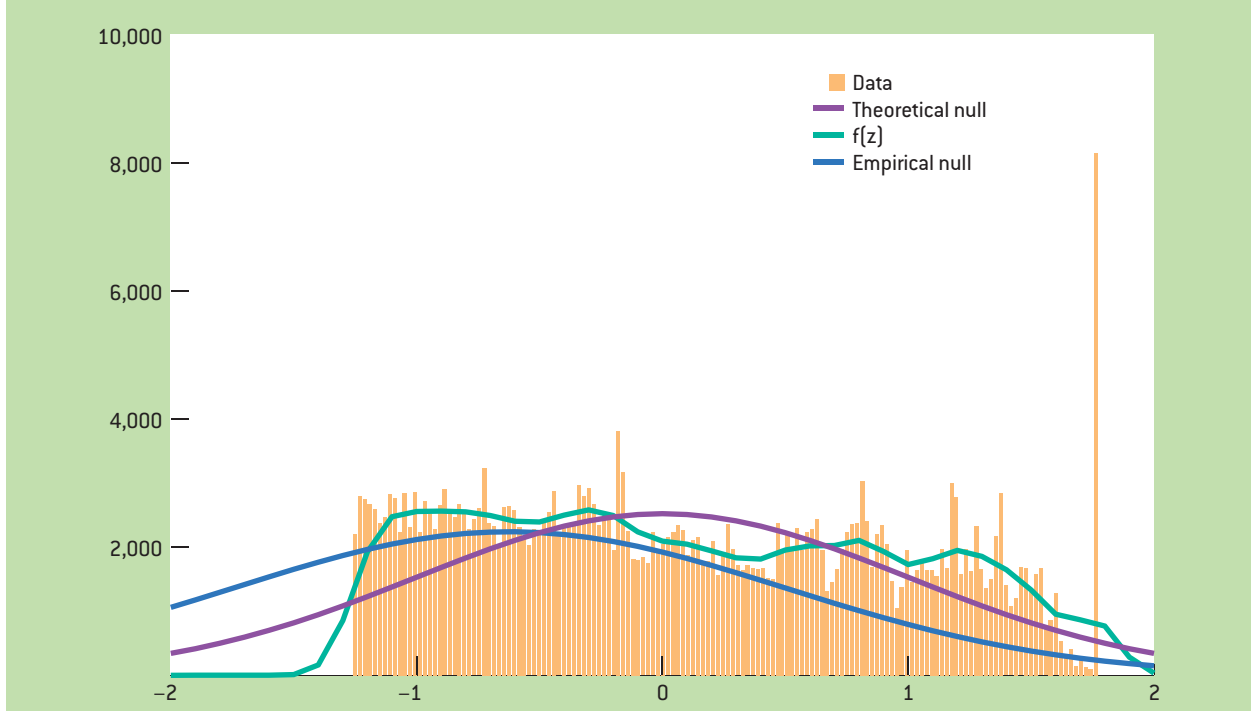**Window 2 (hours 7 to 16) Histogram of Computed $\mathcal{E}_{avg}$**

Data
Theoretical null
f(z)
Empirical null

FIGURE 3

Window 4 (hours 28 to 43) Histogram of Computed $\mathcal{E}_{avg}$

Data
Theoretical null
f(z)
Empirical null

timeline of symptoms, how is it to be interpreted in the context of the previously accumulated knowledge? This question is closely related to the questions about the relationship between type-level (general properties such as "smoking causes cancer") and token-level (particular cases such as "Bob's smoking caused *his* cancer") causalities. Second, when we have two independently developed etiologies from two different sets of patients, how are the inconsistencies to be reconciled? Similar situations arise naturally when we try to combine hypotheses and conclusions about disease etiologies appearing in the literature with the type-level causalities inferred from patient data.

Finally, while our focus here is on systems biology and biomedical applications, undoubtedly, our readers have already correctly surmised that these ideas are much more widely applicable: to neuroscience, finance, politics, Internet social networks, and much more. Indeed, in 2019, Mrs. Samsa uses systems based on these ideas not just to understand her chronic fatigue, but also to manage her finances, carry out her day-to-day social responsibilities, form an opinion about a political candidate, and find the best-suited news blogs—choosing her canaries in the news coalmine. Q

REFERENCES

1. Benjamini, Y., Yekutieli, D. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 29(4):1165–1188.
2. Bozdech, Z., Llinás, M., Pulliam, B.L., Wong, E.D., Zhu, J., et al. 2003. The Transcriptome of the Intraerythrocytic Developmental Cycle of Plasmodium falciparum. *PLoS Biol* 1(1):e5.
3. Clarke, E.M., Grumberg, O., Peled, D.A. 1999. *Model Checking*. Cambridge, MA: MIT Press.
4. Dowe, P. 2000. *Physical Causation*. Cambridge University Press.
5. Dudoit, S., Shaffer, J.P., Boldrick, J.C. 2003. Multiple hypothesis testing in microarray experiments. *Statistical Science* 18(1):71–103.
6. Eells, E. 1991. *Probabilistic Causality*. Cambridge University Press.
7. Efron, B. 2004. Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *Journal of the American Statistical Association* 99(465):96–105.
8. Efron, B., Tibshirani, R. 2002. Empirical Bayes methods and false discovery rates for microarrays. *Genetic Epidemiology*, 23(1):70–86.
9. Friedman, N., Linial, M., Nachman, I., Pe'er, D. 2000. Using Bayesian networks to analyze expression data. *Journal of Computational Biology*, 7(3-4):601–620.
10. Friedman, N., Murphy, K., Russell, S. 1998. Learning the structure of dynamic probabilistic networks. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI'98)*: 139–147.
11. Hansson, H., Jonsson, B. 1994. A logic for reasoning about time and reliability. *Formal Aspects of Computing* 6(5):512–535.
12. Heckerman, D. 1995. A Bayesian approach to learning causal networks. Technical Report MSR-TR-95-04 (March).
13. Jin, J., Cai, T. 2006. Estimating the Null and the proportion of non-Null effects in large-scale multiple comparisons. *Journal of the American Statistical Association*, 102:495–506.
14. Kitano, H. 2004. Biological robustness. *Nature Reviews Genetics*, 5(11):826–837.
15. Kleinberg, S. Casey, K., Mishra, B. 2007. Systems biology via redescription and ontologies (i): finding phase changes with applications to malaria temporal data. *Systems and Synthetic Biology* 1(4):197–205.

16. Langmead, C.J. 2008. Towards inference and learning in dynamic Bayesian networks using generalized evidence. Technical Report CMU-CS-08-151, Carnegie Mellon University.
17. Lewis, D. 1973. Causation. *The Journal of Philosophy* 70(17):556–567.
18. Mitrofanova, A., Kleinberg, S., Carlton, J., Kasif, S., Mishra, B. 2008. Systems biology via redescription and ontologies (iii): Protein classification using malaria parasite's temporal transcriptomic profiles. In *BIBM*, Philadelphia, PA (November).
19. Murphy, K. Mian, S. 1999. Modeling gene expression data using dynamic Bayesian networks. Technical report, University of California, Berkeley, CA.
20. Opgen-Rhein, R., Strimmer, K. 2007. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Systems Biology* 1(1):37.
21. Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press.
22. Salmon, W.C. 1980. Causality: production and propagation. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1980:49–69.
23. Spirtes, P., Glymour, C., Scheines, R. 2000. *Causation, Prediction, and Search*. MIT Press.
24. Spirtes, P., Glymour, C., Scheines, R., Kauffman, S., Aimale, V., Wimberly, F. 2001. Constructing Bayesian network models of gene expression networks from microarray data. *Proceedings of the Atlantic Symposium on Computational Biology, Genome Information Systems and Technology*.
25. Storey, J.D., Tibshirani, R. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440.
26. Suppes, P. 1970. *A probabilistic theory of causality*. North-Holland.

LOVE IT, HATE IT? LET US KNOW

feedback@queue.acm.org

SAMANTHA KLEINBERG (samantha@cs.nyu.edu) is a Ph.D. candidate in computer science at New York University. She focuses on developing methods for understanding systems, such as humans and financial markets, which are observed over time.

BUD MISHRA (mishra@nyu.edu) is a professor of computer science, mathematics, and cell biology at New York University's Courant Institute and School of Medicine. He is also a visiting scholar at the Cold Spring Harbor Laboratory and a fellow of both IEEE and ACM.