Is Cancer Solvable? Towards Efficient and Ethical Biomedical Science

Jeff Shrager, Mark Shapiro, and William Hoos

Billions of dollars are spent annually on cancer research. Yet, despite some progress in some cancers, the majority of patients with advanced cancer still have few good treatment options. In recent years, the concept of biomarker-guided treatment has swept through cancer research. But despite some promising results, the hoped-for benefits to patients have not been widely realized.

One reason that research into biomarker-guided treatments is slow is that it is hard to find good treatments for cancer to begin with, and harder still to associate them with biomarkers. But the limited number of useful treatments is not the root of the problem. Indeed, there are thousands of potential treatments in the pharmaceutical pipelines, and thousands of potentially-useful combinations of approved therapies that have yet to be tested. In fact, as we shall see, searching for targeted treatments — that is, treatments that work only in a narrow molecular subtype of the disease — is actually fundamental to why progress for cancer biomarkers is so challenging.

In this essay, we argue that the major obstacle to discovering new, effective biomarker guided treatments is what computer scientists call "The Curse of Dimensionality."[1] We begin, in the next section, by explaining what the curse is, and how the structure of biomarker-guided cancer treatment leads us into this situation. Then we chart a course that utilizes untapped resources to minimize the impact of the curse, if not entirely lift it.

Science as Search and The Curse of Dimensionality

Computer and cognitive scientists have long thought of problem-solving in terms of the search for good solutions in a potentially very large space of possible solutions.[2] More recently, this way of thinking about research and problem-solving has been generalized to other areas of science, such as astronomy, physics, chemistry, and biology.[3] When applied to medical research, the relevant search is through the space of all possible decision rules that relate particular observations (e.g., a patient's biomarkers) to particular treatments that most effectively alleviate pain and suffering.[4] How quickly we can search this space for highly effective treatments, guided by specific biomarkers, depends upon how large the space is, and upon the efficiency of our search strategy.

Jeff Shrager, Ph.D., *is affiliated with xCures, Inc. and the Symbolic Systems Program, Stanford University.* **Mark Shapiro, M.A., M.B.A.**, *is affiliated with xCures, Inc. and Pharma Initiatives Consulting, LLC.* **William Hoos, M.S., M.B.A.**, *is affiliated with xCures, Inc.*

The Journal of Law, Medicine & Ethics, 47 (2019): 362-368. © 2019 The Author(s)
DOI: 10.1177/1073110519876164

Classically, this search space was not considered to be very large, so a fairly simplistic search strategy based upon large clinical trials was effective. At the peak of the success of classical clinical trials, say 25 years ago, cancer was thought of as what might be called a "10 by 10" disease: There were roughly 10 types of cancer, corresponding to tissue of origin, crossed with roughly 10 types of chemotherapy. This way of thinking is represented by the 100-cell matrix, shown on the left in Figure 1. Each cell in this matrix represents a biomarker-treatment relationship which needs to be tested, with the biomarkers here being the tissue of origin of the cancer, and some other high-level features such as stage and node involvement. For example, non-small cell lung cancer might be a column in this matrix, and the search for the right treatment would involve conducting large, randomized controlled trials for each of the 10 chemotherapies. The roughly 200,000 lung cancer patients each year would provide roughly 10,000 patients per matrix cell (in the lung cancer column), which is plenty of patients to run several large clinical trials of each proposed therapy.

> [W]e argue that the major obstacle to discovering new, effective biomarker guided treatments is what computer scientists call "The Curse of Dimensionality." We begin...by explaining what the curse is, and how the structure of biomarker-guided cancer treatment leads us into this situation. Then we chart a course that utilizes untapped resources to minimize the impact of the curse, if not entirely lift it.

However, as our understanding of cancer has evolved — recognizing that there are potentially thousands of molecular biomarkers that influence whether a treatment will be effective — we are now facing a very different kind of problem. The space of possible treatment decisions is enormously larger, represented by the matrix on the right in Figure 1. With thousands of molecular features, leading to tens of thousands of combinatorial subtypes, and hundreds of plausible combination therapies, there may be many millions of treatment-decision rules (i.e., matrix cells) that have to be tested. Yet, because (thankfully) there are only about a million advanced cancer patients in the US each year, each cell in this matrix will have, on average, approximately *zero* observations.

The number of features that determine the size of this search space is called the "dimensionality" of the data, and the number of independent observations is sometimes referred to as the sample size, or "n" of a study. As a result of the combinatorial structure of the problem, each new feature grows the size of the search space exponentially. This is what computer scientists call "The Curse of Dimensionality"; there will simply never be enough patients and clinical trials to fully explore this space and identify the optimal treatments in the traditional way. In fact, even modern adaptive trials, which can potentially test many decision rules, are impotent in the face of this challenge. Therefore, if we are to make progress in identifying effective biomarker-guided treatment strategies, a new, more efficient search strategy is needed.

AI and Big Data to the Rescue?

Now you may be thinking: "Perhaps 'Big Data' and artificial intelligence (AI) can solve this problem. Isn't this problem exactly what those tools are supposed to help us do?" Unfortunately, Big Data and AI cannot solve this problem. Let us take a short (self-)drive through some of the details of modern AI, examine why it is so successful in some settings, and then see why cancer research is not analogous to these settings.

The settings in which modern AI has been successful typically share one or more of the following five properties:

1. There are well-defined criteria of success, and the signal of success or failure is rapidly available;
2. Data are cheap and plentiful, so that there is a very large amount of data, relative to the size of the search space;
3. There are expert teachers;
4. It is easy and cheap to run experiments;
5. There are highly veridical simulators based upon detailed (e.g., formal, mathematical) models of the underlying physical processes.

These five properties are closely inter-related. For instance, good simulations and models can be used

https://doi.org/10.1177/1073110519876164 Published online by Cambridge University Press

Figure 1

Figure 1. Left: In the "pre-OMIC" era we classified cancers by a handful of features (columns), mostly tissue of origin and stage, crossed with a small number of possible chemotherapeutic treatments (rows), under a small number of protocols. With order 10,000 patients per cell, large trials are efficient. Right: In the present "molecular" era, the feature space is hugely expanded, including molecular and many other observables, and there is a likewise huge combinatorial explosion in treatment number and complexity. As a result there are ~*zero* patients per cell!

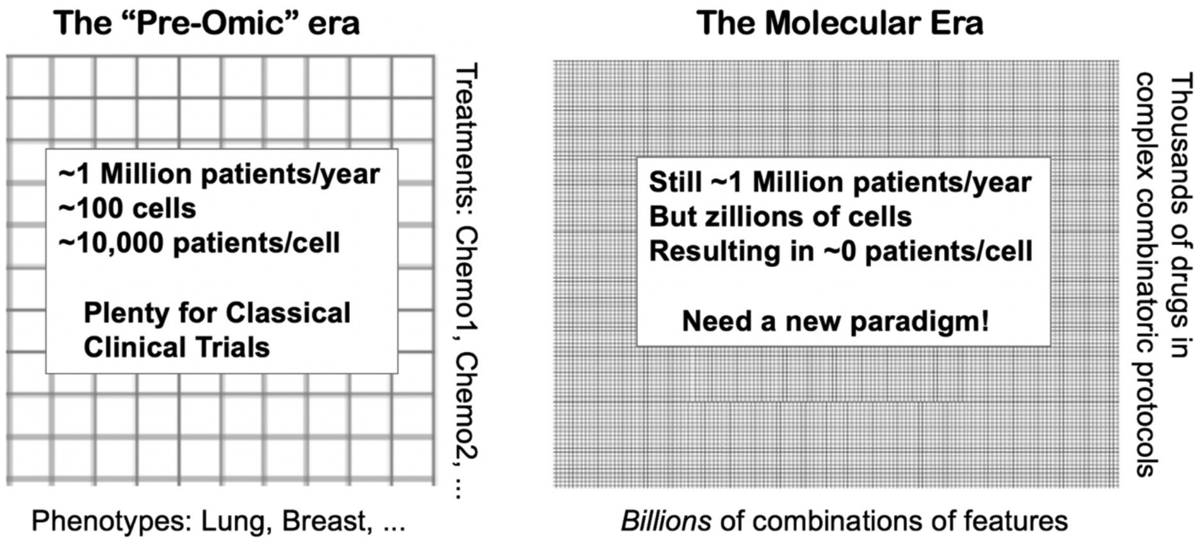
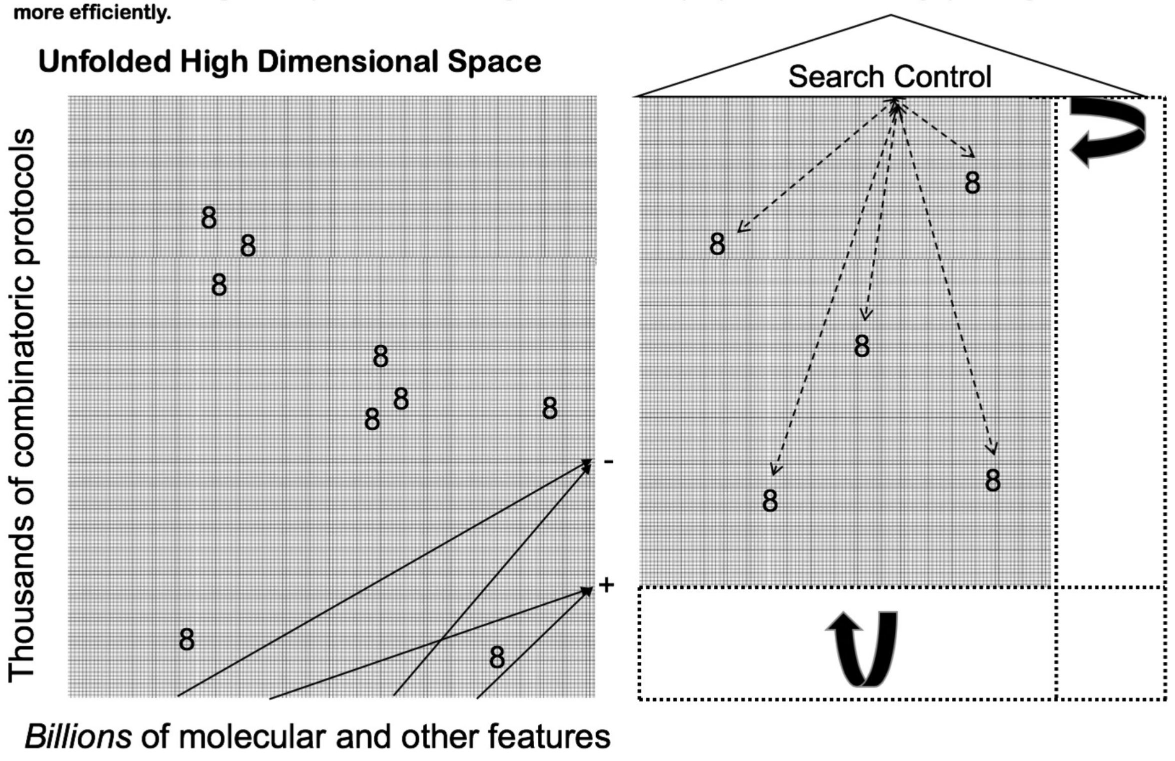

The "Pre-Omic" era
~1 Million patients/year
~100 cells
~10,000 patients/cell
Plenty for Classical Clinical Trials
Treatments: Chemo1, Chemo2, ...
Phenotypes: Lung, Breast, ...
The Molecular Era
Still ~1 Million patients/year
But zillions of cells
Resulting in ~0 patients/cell
Need a new paradigm!
Thousands of drugs in complex combinatoric protocols
Billions of combinations of features

Figure 2

Figure 2. Left: Treatment Rationales (diagonal lines – one positive (+) and one negative (-) TR is shown, each having two observable "inputs" along the base), as well as other sorts of knowledge and data, provide clues that let us hypothesize "folds" (right) that reduce the size of the space. This makes search (ants as "8"s) more tractable. In addition, coordinating search (ants communicating with "controller") improves search rate by spreading the search more efficiently.


Unfolded High Dimensional Space
Thousands of combinatoric protocols
Billions of molecular and other features
8
8
8
8
8
8
8
8
8
-
+
Search Control
8
8
8
8
8

The Journal of Law, Medicine & Ethics, 47 (2019): 362-368. © 2019 The Author(s)

https://doi.org/10.1177/1073110519876164 Published online by Cambridge University Press

to create simulated data, to simulate a teacher, to run simulated experiments, and — as we will describe later — to analytically reduce the size of the search space. Games such as chess and Go, and even domains as seemingly complex as self-driving cars, have all or most of the above properties: Go and chess have been played by humans for hundreds or thousands of years. The rules, and what counts as a win or loss, are completely clear, we have a plethora of expert teachers, and we can build perfect simulators. And, for better or worse, pretty much every teenager drives — or can learn to drive — acceptably well.

We therefore have both an existence proof that Go, chess, and driving can be solved — at least to a nominal level — and we know that we can use human players and drivers to train the self-driving cars and Go/chess computers. Moreover, we can build very good simulators for games because we completely understand, and can perfectly model those problem settings. This is only slightly less the case for driving, where there is a significant social aspect, but the physics of driving can at least be modeled in near-perfect detail, enabling near-perfect simulations.

The domain of medicine, by contrast, has almost none of the above five properties: Medical experiments are extremely costly and we lack good treatments for most diseases, and correspondingly lack expert guidance. Indeed, unlike games and puzzles, or even self-driving cars, it is unclear whether some medical problems have solutions at all. There are no existence proofs for most of medicine. You can't just teach your robot doctor to cure cancer by observing good doctors curing cancer, because there are no such doctors and cures. Moreover, there may well not be a cure for cancer at all.

We also cannot create near perfect simulations of disease. In fact, we struggle to create even good simulations for disease, except for certain simple functions, such as those for which we have dynamical models, such as drug metabolism. The immune-system, which is a critical participant in the preponderance of human disease, is of the order of magnitude of complexity of the human brain, and is similarly a self-modifying machine, so that to model someone's immune profile in detail is similar in difficulty to modeling their entire brain, including all of their learned skills and memories. This is well beyond our current, or even our foreseeable simulation capabilities.

In some fields, experiments are extremely expensive — planetary science, for example, might pay millions of dollars to recover a tiny amount of asteroid dust. Yet medical experiments are in some ways worse. Beyond the in vitro and in silico experiments, medical experiments are both extremely expensive and ethically fraught, both in terms of privacy and because data from in vivo and in-patient experiments can subject the participants to extreme pain and suffering, perhaps even death. For example, a phase 2 cancer clinical trial may need 100 to 500 patients, can cost tens of millions of dollars, and can take years to complete.[5] Consider, by contrast, that Facebook ran an experiment involving almost 700,000 unpaid subjects in the space of a single week.[6] This volume and efficiency of data generation in Facebook's experiment is what is usually referred to as "Big Data." Data in medicine is nothing like this.[7]

Global Cumulative Treatment Analysis

One might, at this point, despair that medicine will ever succeed. We are cursed by dimensionality, with no possibility of simulation, very small amounts of extremely expensive, very high dimensionality data whose gathering may well harm or kill people or animals, and not even an existence proof, with rare exception, that medicine works at all!

But there remain untapped resources that can help us climb out of this admittedly very deep hole. We will describe two of these: (1) Treatment Rationales and (2) System-wide Coordination.

Dimensionality Origami: Folding the Map and the Special Role of Treatment Rationales

One way to make a space easier to search is to make it smaller. Think of the matrix like an old-style paper map — one of those giant pieces of paper with well-worn creases at various places, which, if you could just figure it all out, folds into a conveniently glove-compartment-sized package. If you imagine those creases in our matrix (depicted on the right in Figure 2), each crease allows us to bring together some columns and/or rows, and thereby reduces the overall dimensionality of the space, hence reducing the size of the search space, and improving the efficiency of any search method.

But how are we to discover these "creases" in the map/matrix? The "big data" way to do this is to empirically observe correlations between features. But, as discussed above, this requires much more data than is available, or is likely to become available. Another approach is to use prior knowledge such as formal models, or even partial models. For example, reducing thousands of ACGT base sequences into genes enormously reduces the 3 billion bases of DNA sequence information to around 20,000 features, and this can be reduced even further by recording only aberrations from a standard genome, or by using pathway models, such as BioCyc[8] to relate the genes to one another via their associated protein-catalyzed reactions.

https://doi.org/10.1177/1073110519876164 Published online by Cambridge University Press

Unfortunately, highly veridical biochemical models of mammalian processes, and especially about abnormalities in these processes, are rare.[9] Sweetnam, et al.[10] proposed to augment the sparse existing data and models through the collection of what they termed "Treatment Rationales" (TRs). A TR is an explanation for why a particular course of action (e.g., test, treatment, watchful waiting, etc.) was either recommended or rejected.[11] TRs are distinct from, and complementary to the many other, already duly recorded aspects of treatment such as case history, test results, treatment hypotheses, and outcomes. All of these are critically important to the progress of the search that is biomedicine, but having TRs which represent the explanations underlying the treatment hypotheses, is, in our view, a critical overlooked resource, given the dearth of good biological models. These TRs represent expert knowledge, contextualized by real cases, that help to reduce the dimensionality of the problem by providing hints as to how to fold the space (Fig. 2, left).[12]

The capture and use of TRs also honors both patients' and physicians' knowledge, and the whole process honors their decision authority. Not only do we encourage complete autonomy, but we want to know the reason for physicians' and patients' decisions because these almost certainly provide valuable insights into the domain.

Improving Search Efficiency via Coordination

Reducing the size of the search space is one way to improve search efficiency, another is to improve the search method. Medical research is not being carried out linearly, one patient at a time. There is a great deal of in vitro, in vivo, and in silico work being done in hundreds of labs, and each of these experiments provides some data that can guide our search. Just sticking to patients, there are over a million advanced cancer patients in the US each year, so many of them are being treated at the same time, and large subsets may be being treated for what is, by hypothesis, the same indication. This should afford us over a million in-patient experimental opportunities every year — one for each patient/treatment experience (perhaps several for each, if they undergo several lines of treatment). Unfortunately, in order to take advantage of these opportunities, we need to efficiently coordinate what is done with each patient across the whole medical system, efficiently gather all the data, efficiently integrate it, and efficiently get it back to the front lines where each next patient appears at the door of each next oncologist's office. But medicine is poorly organized to operate this sort of system-wide coordination, depending instead upon results slowly appearing in journals, and upon very slow moving and roughly reasoned coordination through either industrial or government funding processes, which themselves rely upon inefficient peer review processes.

Remarking on the inefficiency of the current system in the face of the vastness of the search problem, Hey and Kesselheim[13] proposed that funding agencies could award responsibility for exploring specific regions of the search space through their grants. However, as Shrager[14] observed, bureaucratic funding agencies cannot react quickly enough to facilitate efficient search of a space this large. Shrager suggested instead that researchers utilize electronic networks and adaptive algorithms to rapidly and adaptively refocus their coordinated research efforts immediately as results become available. Shrager[15] likened this to a "giant distributed robotic discovery system" or biomedical "air traffic control system," or, more precisely, a global adaptive prospective trial, which Shrager called "Global Cumulative Treatment Analysis" (GCTA).[16] Under internet-operated GCTA coordination, all patients would be monitored and treated based on the best available knowledge. When there is uncertainty about whether a treatment will be beneficial, the decision algorithm (be it computational, human, or more likely a combination) assigns treatments to patients in accord with a decision algorithm that incorporates

Current medical research tries to minimize risk to the individual while maximizing benefit to society. Yet, in traditional research, no attempt is made to calculate the value to society of what may be gained by research. In contrast, GCTA seeks to maximize both individual and societal benefit based upon explicit quantification for each. One quantitative measure of societal benefit is information gain, which enables the prioritization of one option versus another at a specific moment in time.

https://doi.org/10.1177/1073110519876164 Published online by Cambridge University Press

patient utility to define a set of equipoise options, and then calculates system-wide information-gain among that set of options. Importantly, because the equipoise set is created *before* the information-gain calculation, the welfare of the individual patient is not subjugated to the needs of society. The parameters of this algorithm would be continuously updated, rapidly integrating response information as soon as it becomes available, and adjusting the course of the system-wide experiment as soon as statistically possible.

Just as in the real air traffic control system (ATC), Shrager's imagined biomedical ATC would be a human-machine system wherein decisions are made locally, and autonomously, but based upon guidance from the "controllers," who have system-wide information and can algorithmically guide this always-on, global, many-armed adaptive experiment. The GCTA process is analogous to the way that granting agencies operate, as envisioned by Hey and Kesselheim, but operating at a much faster pace, and using computational tools that can take advantage of the data in real time in order to guide the search by "vectoring" each next patient in accord with a balance between what is best for that person, and what is best for the system as a whole (i.e., what gains the most information).

Current medical research tries to minimize risk to the individual while maximizing benefit to society. Yet, in traditional research, no attempt is made to calculate the value to society of what may be gained by research. In contrast, GCTA seeks to maximize both individual and societal benefit based upon explicit quantification for each. One quantitative measure of societal benefit is information gain, which enables the prioritization of one option versus another at a specific moment in time.

Conclusion

No invisible hand can hold a half-a-million pounds of 747 in the air, nor keep a hundred thousand commercial flights a day safely separated from one another while taking off and landing more-or-less on schedule. These are large scale engineering problems that have been mastered through large-scale engineering solutions. Efficiently searching a space as large as that faced by medicine, especially in the face of limited, expensive, and ethically fraught data, and the near impossibility of simulation, is a large-scale engineering problem that must be similarly addressed. GCTA is one large-scale engineering approach to the wicked problem of the search for effective treatments for cancer.

At Cancer Commons and xCures we have begun taking baby steps towards GCTA. One such step is the "XCELSIOR" trial (NCT03793088), which is currently accruing advanced cancer patients with a focus on glioma and pancreatic cancer. Under the trial's "perpetual protocol," each patient's case is reviewed by a "virtual tumor board," supported by a platform designed to eventually deploy the human-plus-AI decision-making, analogous to air traffic control, envisioned by GCTA. While we are developing this AI, the Virtual Tumor Boards are populated by natural, not artificial intelligences. (Baby steps!)

Coordinated teams of experts and algorithms enable the air traffic control system to efficiently and safely fly millions of passengers all over the world every day. Similarly, teams of experts and algorithms will enable GTCA to efficiently and safely coordinate the thousands of treatment decisions made every day across the whole medical system, in theory delivering the best possible treatment to every patient while optimally searching the vast space that is the realm of medical research for effective treatments, and sometimes even cures.

No part of GCTA is free from significant theoretical and practical challenges. Yet, to paraphrase a famous philosopher: We are pursuing GCTA not because it is easy, but because it is hard; because it will efficiently organize our energies, skills, and knowledge, and because it is a challenge that we are unwilling to postpone.

Note

This work has been supported by xCures, Cancer Commons, and CommerceNet. Dr. Shrager and Mr. Hoos report a patent pending.

Acknowledgements

Marty Tenenbaum has been instrumental in the conception and development of GCTA. In addition, we are especially indebted to (alphabetically) Asher Wassterman, Robert Bartsch, Kevin Knopf, Nathaniel Knopf, Michael Krauthammer, Lawrence Marton, Allan Schiffman, and Srikar Srinath. Spencer Hey has contributed both conceptually, and by improving this paper.

References

1. Wikipedia, "Curse of Dimensionality," *available at* <https://en.wikipedia.org/wiki/Curse_of_dimensionality> (last visited July 1, 2019).
2. A. Newell and H.A. Simon, "Computer Science as Empirical Inquiry: Symbols and Search," *Communications of the ACM* 19, no. 3 (1976): 113-126; A. Gardner, "Search: An Overview," *AI Magazine* 2, no. 1 (1981): 2-23.
3. H.A. Simon, P. Langley, and G.L. Bradshaw, "Scientific Discovery as Problem Solving," *Synthese* 47 (1981): 1-27; J. Shrager and P. Langley, Eds., *Computational Models of Scientific Discovery and Theory Formation* (San Mateo, CA: Morgan Kaufmann, 1990).
4. "Treatments" are, in reality, usually complex treatment plans, including activities such as cycles of drug infusion, monitoring, additional tests, and so on. Here we will simply call these collectively "treatments."
5. Pharmaceutical Research and Manufacturers of America (PhRMA), "Biopharmaceutical Industry-Sponsored Clinical Trials: Impact on State Economies," 2015; J. Martinez "Driving Drug Innovation and Market Access: Part 1-Clinical Trial Cost Breakdown," 2016, *available at* <https://www.

https://doi.org/10.1177/1073110519876164 Published online by Cambridge University Press

centerpointclinicalservices.com/blog-posts/driving-drive-drug-innovation-and-market-access-part-1-clinical-trial-cost-breakdown/> (last visited July 1, 2019).

6. A.D.I. Kramer, J.E. Guillory, and J.T. Hancock, "Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks," *PNAS* 111, no. 24 (2014): 8788-8790.
7. Two senses of the term "big" are often conflated. The sense of "big data" where, for example, Google and Facebook have enormous datasets, might more accurately be called "tall, narrow data," wherein there are many independent observations of a few features. Put in terms of dimensionality and sample size, from earlier in the paper, there are many independent samples ("n") for relatively few dimensions of feature space. This "tall narrow data" might be more correctly called "large n, low (or moderate) dimensionality data." Medical data, especially at the molecular level, might better be called "short wide data," with relatively few independent observations over a very large number of features, or, in the above parlance: "small n, high dimensionality data." The "big data" problems that have seen success are the tall narrow ones. The short wide ones remain out of reach of current technology.
8. P.D. Karp, R. Billington, R. Caspi, C.A. Fulcher, M. Latendresse, A. Kothari, I.M. Keseler, M. Krummenacker, P.E. Midford, Q. Ong, W.K. Ong, S.M. Paley, and P. Subhraveti, "The BioCyc Collection of Microbial Genomes and Metabolic Pathways," *Briefings in Bioinformatics* (2017): bbx085, doi:10.1093/bib/bbx085.
9. J. Shrager, "The Fiction of Function," *Bioinformatics* 19, no. 15 (2003): 1934–1936, doi:10.1093/bioinformatics/btg277.
10. C. Sweetnam, S. Mocellin, M. Krauthammer, N. Knopf, R. Baertsch, and J. Shrager, "Prototyping a Precision Oncology 3.0 Rapid Learning Platform," *BMC* Bioinformatics 19 (2018): 341, doi:10.1186/s12859-018-2374-0.
11. Sweetnam et al. note that TRs should be captured not only for the recommended treatments (tests, etc.), but also for those actions that are considered but rejected, because they are incorrect, undesirable, or infeasible. These "contra treatment rationales" ("contra-TRs") can carry as much, or in some cases more, information than the rationale supporting the final recommendation; often the final recommendation is a safe or possible choice, whereas physicians might like to do something that may be more effective if practical barriers, such as cost or side effects, could be surmounted. Contra TRs may also represent new treatment hypotheses, possibly worthy of testing.
12. Even though reasoning based on these partial explanations would likely be imperfect, these can serve as hypothetical "creases" where the space could fold, as depicted in Figure 2. (Unless they are analytic [for example, based upon mathematical formulae], such hypothetical folds need to be treated with statistical care regarding Type I Error [False Discovery]). The present hypothesis is that such hints do more to reduce the dimensionality of the problem than they do to increase the false discovery rate.
13. S.P. Hey and A.S. Kesselheim, "Countering Imprecision in Precision Medicine," *Science*, July 29, 2016.
14. J. Shrager, "Precision Medicine: Fantasy Meets Reality," Letters, *Science* 353, no. 6305 (2016).
15. *Id.*
16. J. Shrager, "Theoretical Issues for Global Cumulative Treatment Analysis (GCTA)," arXiv: 1308.1066v1 [stat.AP] (2013); J. Shrager and J.M. Tenenbaum, "Rapid Learning Precision Oncology," *Nature Reviews Clinical Oncology* 11 (2014): 109-118.

The Journal of Law, Medicine & Ethics, 47 (2019): 362-368. © 2019 The Author(s)