Rapid learning for precision oncology

Jeff Shrager and Jay M. Tenenbaum

Abstract | The emerging paradigm of Precision Oncology 3.0 uses panomics and sophisticated methods of statistical reverse engineering to hypothesize the putative networks that drive a given patient's tumour, and to attack these drivers with combinations of targeted therapies. Here, we review a paradigm termed Rapid Learning Precision Oncology wherein every treatment event is considered as a probe that simultaneously treats the patient and provides an opportunity to validate and refine the models on which the treatment decisions are based. Implementation of Rapid Learning Precision Oncology requires overcoming a host of challenges that include developing analytical tools, capturing the information from each patient encounter and rapidly extrapolating it to other patients, coordinating many patient encounters to efficiently search for effective treatments, and overcoming economic, social and structural impediments, such as obtaining access to, and reimbursement for, investigational drugs.

Shrager, J. & Tenenbaum, J. M. *Nat. Rev. Clin. Oncol*. advance online publication 21 January 2014; doi:10.1038/nrclinonc.2013.244

Introduction

Panomic analyses (such as genomic, transcriptomic, proteomic, metabolomic, etc) demonstrate that, although cancer genomes may harbour thousands of aberrations,[1,2] the networks of molecular pathways that drive an individual patient's tumour can be elucidated.[3–6] The emerging paradigm of Precision Oncology seeks to determine tumour-driving networks that function in a particular patient's tumour, and to then design a rational combination therapy—selected from the rapidly growing arsenal of targeted drugs—that will ameliorate the effect of the aberrations in that particular patient's tumour. This approach has the advantage that narrowly targeted therapies may have fewer side effects, and may be more effective than broad cytotoxic therapies. However, elucidating the driver networks and creating effective combination therapies is cutting-edge biomedical science, and it is critical to do this correctly, otherwise the tumour might escape the treatment altogether. Moreover, even if the correct targets are determined, there may be no available treatments for some targets.

In this Review, we briefly describe what we call Precision Oncology 3.0, and then focus on the challenge of efficiently searching for the optimal combination therapy recognizing that cancer constitutes thousands of functional subtypes that can be treated with hundreds of targeted drugs. We describe approaches to coordinate this search across all patient encounters (Box 1), and to capture what is learned from each such encounter and rapidly apply it to other patients. We also discuss approaches to overcome some of the economic, social and structural impediments faced by these efforts, such as obtaining access to, and reimbursement for, investigational drugs.

CommerceNet (**J. Shrager**), Cancer Commons (**J. M. Tenenbaum**), 955-A Alma Street, Palo Alto, CA 94301, USA.

Correspondence to: J. Shrager jshrager@stanford.edu

Competing interests
The authors declare no competing interests.

Precision Oncology 3.0

Borrowing from Web nomenclature, one can roughly distinguish three generations of Precision Oncology. Precision Oncology 1.0, the prevailing standard, involves testing for small numbers of molecular abnormalities that are correlated with drug response in particular tumour types (for example, companion diagnostics [Box 1] for EGFR inhibitors in lung cancer). Precision Oncology 1.0 is almost always constrained by the tissue-of-origin, and other non-molecular characteristics such as microscopic histology. Precision Oncology 2.0 involves examining dozens or potentially hundreds of possible mutational hotspots simultaneously,[7] or sequencing the exomes of several hundred cancer-associated genes,[8] and this approach might sometimes disregard non-molecular characteristics. Whereas the tests and interpretations involved in Precision Oncology 1.0 are narrowly constrained and well defined, Precision Oncology 2.0 requires that laboratories have specialized equipment, perhaps including next-generation sequencing, and imposes a much greater interpretive load on the physician, who may be expected to develop a therapeutic regimen to match a wide range of possible molecular subtypes (Box 1). Because of these requirements, few patients have had the opportunity to take advantage of Precision Oncology 2.0, but with the broad availability of next-generation sequencing and molecular diagnostic service providers to aid in interpretation,[9–11] it is rapidly becoming the standard of care at leading cancer centres worldwide.

Unfortunately, cancer biology is far too complex to be characterized by the mutational status of a few genes.[12] Most cancers seem to arise from collections of mutated and aberrantly regulated genes that collaborate to promote tumour growth.[13] The picture is further muddied by feedback cycles, immune responses, and pharmacokinetics. Moreover, tumours can accumulate further mutations to escape a given treatment. These facts suggest that cancer

Key points

- In Precision Oncology 3.0 sophisticated algorithms analyse panomic data to hypothesize the molecular pathways that drive an individual patient's tumour, and hypothesize personalized treatments, using combinations of narrowly targeted therapies
- At the molecular level, where Precision Oncology 3.0 operates, there are far too many combinations of driver mutations and possible treatments to be efficiently searched by current clinical trial methodologies
- The 'Rapid Learning Precision Oncology' paradigm considers each patient encounter as an experiment, continuously gathering and analysing all the data to inform each subsequent encounter with the same or similar patients
- All patient encounters can be coordinated through a 'Global Cumulative Treatment Analysis' (GCTA) methodology, which chooses treatments according to their continuously updated performance statistics
- The Rapid Learning approach can help to overcome some of the technical and structural barriers facing Precision Oncology 3.0, including the facilitation of the off-label uses of targeted drugs

must be managed by treatment regimens comprising a cocktail of drugs that are changed to challenge the evolving disease. The emerging new generation of Precision Oncology, which we term Precision Oncology 3.0 (Figure 1), uses broad-spectrum panomics and sophisticated network-based statistical reverse engineering (Box 1) methods to hypothesize the putative driver networks for a given patient's tumour. Once these are computed, they are combined with important contextual features (such as the patient's treatment history, status, and preferences, as well as knowledge of available drugs and drug interactions) to hypothesize a treatment plan that attacks these tumour drivers with cocktails of narrowly targeted therapies.[14–16]

In Precision Oncology 3.0, as we define it, initial treatment decisions, monitoring, and subsequent treatment choices are all based on sophisticated molecular analysis of the biochemical and signalling networks involved in the disease (Figure 2). For example, if an initial therapy designed to interrupt a signalling pathway fails to inhibit tumour growth, one would analyse whether the treatment reached its targets, and whether the pathway was blocked as intended. If the patient initially responds but develops resistance, subsequent analysis would be undertaken to reveal how the tumour has evolved at the molecular level to evade the drugs, suggesting new lines of therapy for that patient, and for other patients with similar tumours. Panomic comparison of the tumour before, and following, the development of resistance (possibly with additional intervening biopsies) could characterize the driver networks in detail. Serial monitoring of the patient's tumour could also provide insight into how these networks work by revealing how they dynamically respond to perturbations.[17] By charting the trajectory of a tumour's molecular profile over time, it might be possible to anticipate how a cancer is likely to evolve, and to take proactive steps to block it from doing so. One might also examine the consequences of perturbing tumour networks by combining targeted drugs and cytotoxic therapies.[18] In Precision Oncology 3.0, every treatment event is a probe, simultaneously treating the patient and providing an opportunity to test and improve our molecular understanding of the disease.[3,4,19] Whereas classical clinical trials provide strong evidence for the efficacy and/or effectiveness (or lack thereof) of a small set of treatments in a large set of patients, Precision Oncology 3.0 works in the opposite way, evaluating a wide range of possible treatments in a small cohort of patients, and then aggregating the results over all such experiments to achieve strong evidence. Moreover, by capturing what is learned about each pathway and each drug at each such encounter, the resulting knowledge can be generalized to other drugs or drug combinations, patients, and cancers, enabling learning to proceed rapidly, one patient at a time instead of one trial at a time.

The concept of molecular reverse engineering of driver networks is widely sought after, even if the technology required to do so has only recently become widely available. As early as 2002, computational biologists were engaged in reverse engineering the networks that are active in particular cells by analysing expression data,[20–22] and in the subsequent decade this technology has blossomed to the point where it is practical on a large scale.[3,4,23–25] Examples of drivers of cancers that have been analysed in this way include the reverse engineering of the nuclear receptor TLX oncogenic transcriptional network to identify the transcription factor RUNX1 as a tumour suppressor in T-cell acute lymphoblastic leukaemia (T-ALL),[26] and the determination of a transcriptional module that activates expression of mesenchymal genes in malignant glioma, including the transcription factors C/EBPβ and STAT3 as master regulators of mesenchymal transformation.[14]

Evidence for the ability of Precision Oncology 3.0 to learn from single patients comes from the analysis of exceptional responders (Box 1) in large-scale clinical trials. Two examples were recently reported. The first example involved a rare responder to everolimus (in a trial for advanced bladder cancer) whose tumour harboured an unexpected *TSC1* mutation that was subsequently identified in several of the partial responders in the same study.[27] The second example involves a complete responder in a negative phase I safety trial that tested a chemotherapy drug combined with an experimental drug that interfered with DNA repair. The patient harboured a mutation in *RAD50*, a gene that codes for proteins involved in a DNA repair complex, which is found to be mutated in 4% of all patients with cancer, suggesting that they too might benefit from this 'failed' drug combination.[28]

Investigators at some major cancer centres are beginning to apply Precision Oncology 3.0 in a clinical setting, along with a few commercial vendors. In addition to the examples mentioned above,[3,4,6,7] medical institutions pursuing Precision Oncology 3.0 include (but are not limited to) the Duke Centre for Personalized and Precision Medicine (CPPM),[29] the Institute for Precision Medicine at the Weill Cornell Medical College at New York–Presbyterian Hospital (New York, NY),[30] the MD Anderson Cancer Centre Institute for Personalized Cancer Therapy (Houston, TX),[31] the Centre for Translational Pathology at the University of Michigan (Ann Arbor, MI),[32] and the Personalized Cancer Medicine Program at the Icahn Institute for Genomics and Multiscale Biology at Mount

Box 1 | Key terms

Patient encounter
Any interaction between a patient and a medical institution that leaves a record, for example seeing one's doctor, having a test performed, or undergoing a treatment.

Companion diagnostics
Tests that help physicians make decisions about particular treatments. For example, specific mutations in genes encoding EGFR pathway proteins may predict differential sensitivity to cetuximab.

Molecular subtypes
Types of disease defined based on molecular characteristics, rather than, or perhaps in addition to, tissue of origin and/or histological characteristics. For cancer, molecular subtyping is usually based on genomic aberrations, but some approaches prefer to subtype cancers at the level of malfunctioning biochemical and signalling pathways.

Statistical reverse engineering
A computational method that combines data and knowledge to rank alternative hypotheses according to their statistical likelihood. For example, using expression data to rank different possible signalling pathway hypotheses according to those that are most likely to be malfunctioning in a given tumour.

Exceptional responders
Patients who respond unusually well or unusually poorly to a treatment, with respect to the typical response in a clinical trial. 'Exceptional' is not a well-defined statistical level (contrast: 'statistical significance'), so its meaning must be specified at each application.

Dimensionality (high dimensional)
The dimensionality of data describes the number of different attributes of the observations. For example, the classical description of a tumour by tissue of origin and number of lymph nodes involved is 'low dimensional', versus the panomic description of a tumour, which may have tens or hundreds of thousands of dimensions, which is very 'high dimensional'. Both low and high dimensionality data create challenges in analysis.

Knowledge base
A kind of database that contains 'knowledge' rather than 'data'. The distinction between 'knowledge' and 'data' is philosophically subtle. Roughly speaking, whereas data are usually concrete observations—such as test results—knowledge is usually abstract, possibly putative, facts, such as that the EGFR signalling pathway comprises certain proteins.

Unsupervised hierarchical clustering
A statistical process whereby high dimensionality observations are automatically arranged into clusters (that is, groups or categories) that reveal strong but hidden, and potentially complex, correlations among the observations. For example, tissue of origin is the most common current clinical categorization of cancers, but molecular oncology hypothesizes that with more data along many more dimensions, as is afforded by panomic analysis, a stronger categorization will arise based on aberrations at the genomic or pathway levels.

Software source code control
A development methodology widely used by software engineers to keep track of changes in computer programmes, commonly when there are many software engineers involved in the project. The method usually depends on specialized computer programmes that maintain a central repository of the source code and keep track of all the changes made, who made them, when and why they were made. This software also helps resolve conflicting changes.

Open source engineering
An engineering methodology wherein the implementation details of a mechanism or method, such as engineering plans or computer programmes, are publically revealed, and others are encouraged to examine, validate, and build on them. Compare, for example, the open pre-publication of the methods of a clinical trial, versus the closed algorithmic details of medical device that are considered trade secrets.

Sinai Hospital (New York, NY).[33] The Personalized Cancer Medicine Program at Mount Sinai Hospital is representative as, to date, it has run 20 cases, spanning a variety of cancer types (ovarian, breast, colorectal, glioblastoma, and pancreatic), six of which have yielded potentially clinically actionable findings, such as a relevant clinical trial or biomarker (Eric Schadt, personal communication to J. S.). Examples of commercial venders offering Precision Oncology 3.0 services include GeneKey, Inc. (Palo Alto, CA)[34] and N-of-One, Inc. (Lexington, MA).[35]

Unfortunately, we are still climbing a very steep learning curve. The high variability of cancer, the enormous amount of very complex data delivered by panomic technologies, the large number of targeted therapies under development, and the need for combined regimens, all distributed over a considerable, but distinctly finite number of patients, render cancer, in effect, a large number of rare diseases occupying a very high dimensional space (Box 1), with very few opportunities for action and observation in each subtype. To efficiently search a space of this nature, one needs to capture the learnings from as many patients and treatment experiments as possible in a continuously updated knowledge base (Box 1), and use that knowledge to guide each treatment decision across all patients in a coordinated manner that optimizes the tradeoffs between patient outcomes and knowledge acquisition. We call this process 'Rapid Learning Precision Oncology'.

Rapid learning

In 2007, the Institute of Medicine (IOM) proposed a framework to use clinical data that is routinely collected, to drive scientific discovery; they called this framework 'rapid-learning health care'.[36,37] The IOM envisioned using electronic health data to generate and test hypotheses about treatment effectiveness quickly and inexpensively, especially for patient cohorts typically excluded from clinical trials, such as elderly patients with advanced-stage disease, patients with multiple comorbidities, and patients on concomitant medications. Shortly thereafter, the IOM proposed a similar rapid-learning system for oncology,[38,39] in which several cases were developed to, for example, compare the effectiveness of multidrug combinations, and use the historical experiences of similar patients to help guide treatment choices for a current patient.

As early as 2008 some institutions had launched rapid-learning efforts for Precision Oncology 2.0. For example, the Moffitt Cancer Center (Tampa, FL) launched Total Cancer Care®,[40] creating an infrastructure (including specimen handling standard operating procedures [SOPs], data exchange standards, and patient consents) whereby 17 community hospitals can send tumour specimens and clinical data to a centralized repository, maintained by Moffitt affiliate M2Gen®. In return, the originating hospital receives support from M2Gen® in undertaking collective analyses of stored data and specimens. This network can then be used for various collective translational research functions, such as cohort identification, clinical trial recruiting, and comparative effectiveness studies. For example, Ren *et al.*[41] found 50 truncating mutations in JAK1 in 36 of 635 gynaecological tumours in the Total Cancer Care® (TCC®) tumour bank. Whereas cancer-associated protein tyrosine kinase (PTK) mutations usually confer a gain-of-function, JAK1 deficient cancer cells are defective in IFN-γ-induced LMP2 and

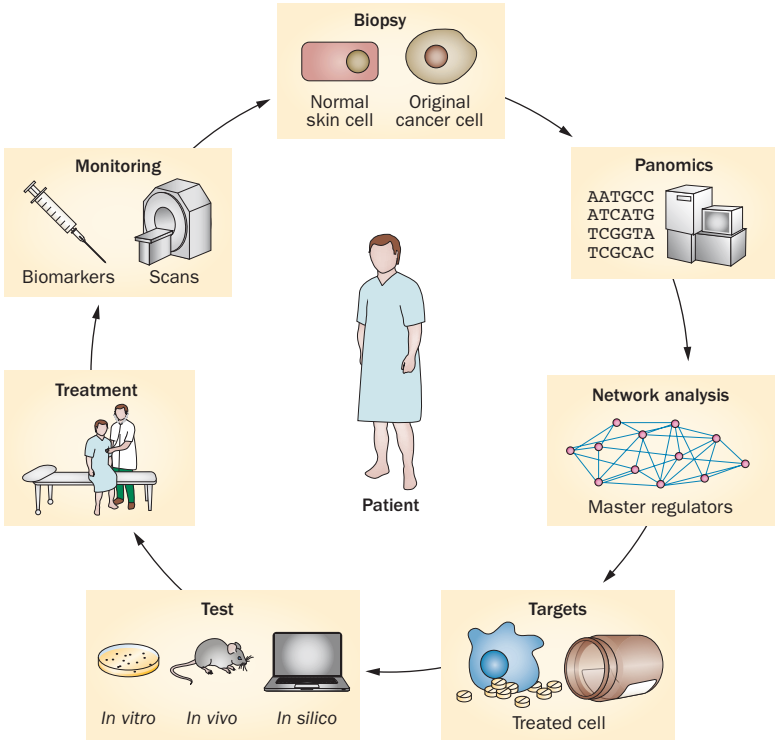

Figure 1 | Precision Oncology 3.0 in outline. Panomic data from tissue samples, obtained from tumours and surrounding healthy tissue, are analysed to produce a list of hypothetical aberrant driver networks. Drug candidates that target specific molecular pathways are selected and validated in models if possible; for example, *in vitro* (in tumour-derived cell lines) or *in vivo* (in mouse models). If the decision is made to move forward with that treatment, the patient is treated and monitored using rapid measures, such as imaging and serum biomarkers. Failure to respond, or disease recurrence, might lead one to choose a different drug or combination based upon a fresh analysis.

TAP1 expression, loss of which inhibits presentation of tumour antigens. These results suggested that recurrent JAK1 truncating mutations could contribute to tumour immune evasion in the studied cancers.

In theory, applying the principles of rapid learning to Precision Oncology 3.0 would provide each patient with the best possible treatment based on the latest knowledge, while efficiently gathering evidence to advance our understanding of cancer mechanisms, molecular subtypes, and therapies. However, its implementation will require overcoming a host of challenges, such as developing computational and analytical tools to distill the data from each patient into an accurate model of the driver networks, capturing the information about these networks under the influence of drug combinations, and rapidly translating this knowledge to support clinical decision making. Other challenges include coordinating the tens of thousands of separate patient encounters in a way that possible molecular subtypes and treatments can be identified efficiently to avoid repeating the same mistakes, while replicating the successes; and overcoming economic, social, and structural impediments such as obtaining early access to investigational new drugs (INDs), and convincing payers (such as insurance companies and national health plans) to cover rational, off-label use of approved drugs.

Several technical approaches can address aspects of these challenges: seeking correlations in existing data (data mining or big data analysis), detailed analysis of particular cases (small data analysis), rapid scientific communications, and coordination through a process termed Global Cumulative Treatment Analysis (GCTA).

Big data analysis

The bioinformatic analysis of large datasets, based on recurring and co-occurring mutations, copy number variations, and aberrant RNA expression levels in patients with similar and different types of cancer, has long been a mainstay of cancer research.[12] Precision Oncology relies heavily on big data results, such as the characterization of canonical pathways and statistical subtypes to understand and treat an individual patient. For example, analyses of the data deposited in The Cancer Genome Atlas have revealed that cancer genomes contain small numbers of recurrent mutations accompanied by much larger numbers of mutations that are rarely or never found in the tumours of other patients.[42] Although recurring mutations are probably involved in disease pathogenesis, rare or unique mutations likely represent a mixture of both driver mutations that collaborate to promote tumour growth, and passenger mutations that are uninvolved in cancer genesis or progression.[43,44] Analysis of large gene-expression array datasets, obtained from matched tumour and normal tissue samples, juxtaposed with these driver mutations, led to the mapping of canonical cancer pathways.[45–47] Statistical learning techniques can infer additional clinically actionable information.[48] For example, unsupervised hierarchical clustering (Box 1) applied to transcriptome data from patients with breast cancer recently revealed at least ten distinct molecular subtypes, which correlate well with observed therapeutic responses.[49,50] These results have the potential to immediately impact the clinical management of breast cancer, which is usually based on just four subtypes: ER-positive, PR-positive, HER2-positive and triple-negative breast cancer.

Small data analysis

The 'big data' initiatives will provide important population analyses, however, one can also learn incrementally from individual patients and small cohorts. Although the data from an individual tumour is generally insufficient to map its pathways *de novo*, there is often enough data to identify which known oncogenic pathways are active, and even to recognize previously unknown perturbations, providing an immediate opportunity to expand our knowledge of the variability of such networks of pathways. The examples regarding the analysis of exceptional responders, described earlier, illustrate the potential of learning from small data. Motivated in part by those results, Deputy Director James Doroshow of the National Cancer Institute (NCI) issued a call at the 2013 American Association for Cancer Research (AACR) Annual Meeting for researchers to identify 100 exceptional responders from failed trials for panomic analysis.[28]

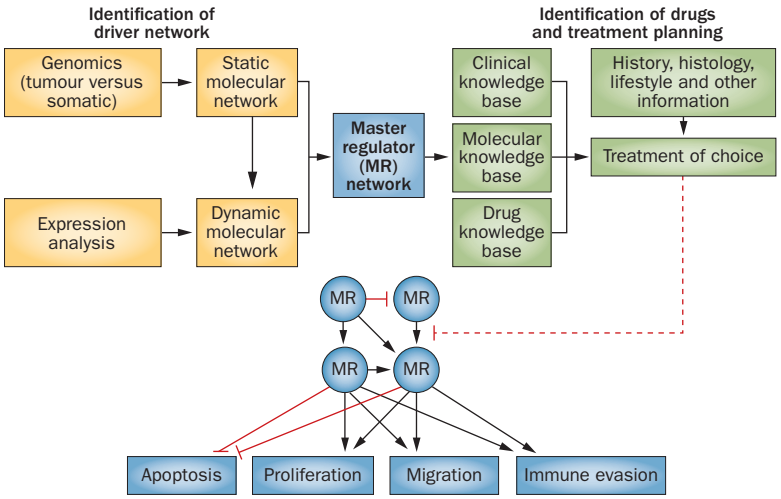

Figure 2 | Precision Oncology 3.0 core algorithms and components. The heart of Precision Oncology 3.0 is driver network analysis and clinical targeting and treatment planning. Driver network analysis identifies key genes, called master regulators,[13] which modulate established cancer hallmarks, such as aberrant proliferation, immune evasion, or circumvention of programmed cell death. Clinical targeting and treatment planning creates treatment hypotheses on the basis of the hypothesized master regulators, combined with clinical and contextual knowledge, such as drug approval status, patient history, drug interaction knowledge, and so on. Arrows in the master regulator network diagram indicate regulatory relationships within the signalling pathways. Dashed arrows indicate putative drugs that modulate identified master regulators.

Rapid communications

Rapid Learning demands a fast and flexible communication channel.[51] A recent study from the National Academy of Sciences warned that "biomedical research information can take years to trickle to doctors and patients, while wasteful health-care expenditures are carried out for treatments that are only effective in specific subgroups".[52] One pioneering effort to facilitate the rapid communications of high quality biomedical information is PLoS Currents,[53] which offers scientists rapid publication in time-sensitive domains such as disasters and influenza, but also in less time-critical areas, such as Huntington's disease. One of these, Evidence on Genomic Tests, is especially close to the spirit of a flexible rapid communications channel for precision oncology. A similar initiative is being developed by Rapid Science Inc. (a non-profit spin-off of Cancer Commons, discussed below), which will publish articles on clinically actionable findings, including knowledge-base updates, case reports, trial results, annotated rare variants, etc. All the contents of these online 'e-journals' are peer reviewed, PubMed indexed, and available under open access licences. In the best case, e-journals such as these will be built on computational platforms inspired by software source code control (Box 1),[54] permitting articles to be version-controlled and revised to incorporate new results. Moreover, such platforms could be integrated with the platforms that maintain the Precision Oncology knowledge base so that when updates are made to the knowledge base, these could be published automatically in the e-journal, and vice versa.

Coordination

To efficiently search a space as large and sparse as the one we are faced with in cancer, every patient encounter should be treated as an information gathering opportunity. However, this is not enough. Treatment decisions must also be coordinated across all patient encounters to avoid unnecessary replication of either positive or negative experiments and to maximize the amount of information obtained from every encounter.

Clinical trials are examples of this sort of coordination, but classical trials are not efficient enough in either speed or breadth to search the vast space of cancer subtypes and treatments. Adaptive trials, one way to improve the efficiency of searching possible molecular subtypes and treatments,[55–60] are now mainstream science. Researchers have recently taken the adaptive trial concept to its natural conclusion, proposing what we term 'Global Cumulative Treatment Analysis' (GCTA; Figure 3),[61–65] wherein decision making, data collection, and data analysis are continuous and integrated, and all available performance data for every rational therapeutic regimen is taken into account to rank treatment options at each decision point for every patient (Figure 3).

Notably, GCTA is neither a big data/data mining approach nor a purely rapid learning approach in the IOM sense, both of which primarily take advantage of the data resulting from mostly uncoordinated treatment events, or from clinical trials. By contrast, the GCTA approach is both global and prospective, explicitly manipulating the treatment rankings offered to a given patient at the point of care by combining all the available information across all patients, to ensure that the appropriate level of experimental variability is injected into the distribution of patients receiving different treatments (Figure 3). This coordination can only be accomplished if there is an effective rapid communication methodology in place, as described above.

Initiatives

CancerLinQ

The most ambitious implementation, to date, of IOM's 2010 rapid learning concept is ASCO's CancerLinQ initiative,[66] which aims to use data from electronic health records to improve quality of care, value, and outcomes in community oncology practices. ASCO expresses their vision for CancerLinQ as follows: "today, we know very little about most patients with cancer—from the molecular characteristics of their tumours to the outcomes of their treatments—because these details are locked away in unconnected electronic and paper records."[67,68] ASCO's vision is to assemble and analyse all of the information in a central knowledge base that will grow continuously. CancerLinQ will "upload clinical data stored in electronic health records (EHRs) […] from patients in multiple practices; aggregate information from EHRs, new clinical trials and published guidelines; identify trends and associations between myriad variables, in order to generate new hypotheses; allow physicians and researchers to evaluate those hypotheses and determine which ones may lead to improved care in real-world settings;

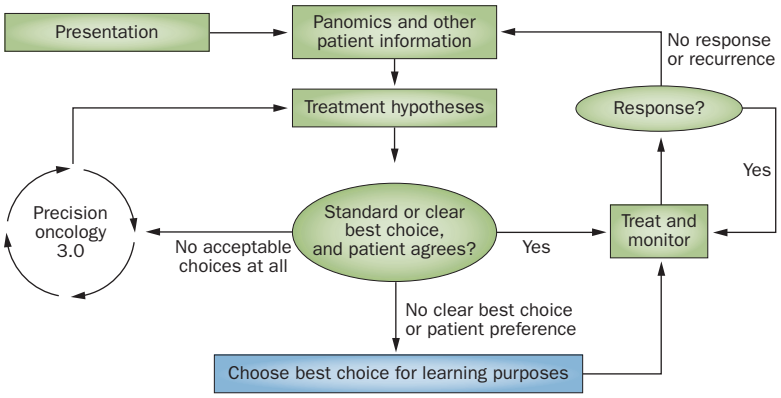

Figure 3 | Global Cumulative Treatment Analysis (GCTA). In current standard practice, a patient presents with certain symptoms, tests (such as a biopsy) are carried out and results are obtained, treatment choices are then ranked, usually on the basis of the treating physician's knowledge and the standard of care, and one of these treatments is chosen. As treatment proceeds, progress is monitored, and either the disease resolves, or recurs, in which case more tests and latter line treatments are needed. If no standard treatment options are available, as is commonplace for advanced-stage cancer, the patient's options may include non-standard treatments, or clinical trials. In Rapid Learning Precision Oncology decision making takes into account all the available performance data over all possible treatment experiences, augmenting the physician's knowledge with systems-biology-based treatment ranking algorithms. Moreover, when there are no acceptable choices at all, one might embark on deep molecular analysis. Discoveries from this process, whether or not they benefit the present patient, are returned to the knowledge base to benefit future patients. When treatment options exist, but there is no clear superior one, the available choices are algorithmically reordered to reflect the information gathering value of each.[61]

and enable clinicians and researchers to quickly apply those conclusions, forming a continuous cycle of learning."[67] In March 2013, ASCO completed a CancerLinQ prototype, encompassing records from more than 100,000 patients with breast cancer, and plans to proceed with a full-scale implementation.[67,68]

Watson

A project with similar goals to CancerLinQ is the collaboration between IBM's Watson group and Memorial Sloan–Kettering Cancer Centre[69,70] (MSKCC). Watson is a machine learning system, developed by IBM, which is being trained to recommend trials and individualized therapies, based on knowledge of clinical oncology obtained from medical text books, journals, clinical trial databases, and many other sources. To date, Watson has analysed and incorporated the knowledge from millions of published articles and other sources. To apply this knowledge, Watson must learn to evaluate the relevance of each bit of information to a patient case, taking into account the strength of evidence of the information (for example, results from a randomized trial are better than a case report, which is better than an animal study, and so on). In addition to reading, Watson is provided with training cases and feedback from experts.

Initial efforts at MSKCC have focused primarily on metastatic NSCLC and early stage breast cancer. Ultimately, the goal is to have Watson serve as the engine for rapid learning, recommending treatment options, and continuously improving these recommendations on the basis of the resulting outcomes. Given the complexity of cancer and the exponential rate at which knowledge is growing, it is inevitable that computational engines (such as Watson), as well as those that operate the GCTA process, will be an essential component of Rapid Learning Precision Oncology.

Cancer Commons

Whereas CancerLinQ is focused on electronic health records and Watson adds broad medical knowledge, the Cancer Commons initiative[64,65] provides a rapid learning infrastructure specifically for Precision Oncology. In Cancer Commons, physicians, scientists, and patients collaborate in rapid learning communities, in which patients are treated in accord with the latest knowledge on molecular subtypes and therapies. A peer-reviewed knowledge base, documents the community's collective understanding of cancer in terms of molecular subtypes, treatments, and trials, with links to relevant papers, data, case reports, news, and other resources.[71–73] The goal of this initiative is to keep this knowledge continually updated—on the basis of each patient's response, as well as all publicly available information—and to rapidly disseminate the updates through e-journals, such as those discussed above.

To close the rapid learning loop, Cancer Commons gathers panomic and clinical data about individual patients from participating oncologists and the patients themselves, in a so-called Donate Your Data (DYD) registry.[74] These data will be used initially to quantify, validate, and refine information in the knowledge base. For example, physicians and patients can compare the actual response rates in a given tumour subtype when choosing among recommended treatments. If patients respond differently to a given treatment, it might be appropriate to split that subtype, corresponding to responders and non-responders, or to add a new subtype to accommodate a previously unseen molecular driver.

Both CancerLinQ and Cancer Commons plan to de-identify all patient data and make it publicly available for research, in the spirit of the proposed 'e-trials'[75] wherein phase II/III trials are replaced by physicians dispensing experimental medicines and gathering response data in a central repository open for analysis by any qualified medical researcher. The response of any patient or group of patients to a drug or treatment could then be compared with those of others in the database who were treated in a different manner or not at all. This could enable many treatment hypotheses that are tested today in investigator-initiated trials, to be studied at a fraction of the time and for a fraction of the cost. For example, exceptional responders could be identified, and their data retrospectively analysed to try and understand why they responded or recurred. If the results are promising, a small cohort of patients with the right subtype could be rapidly recruited through something like the DYD registry to validate the findings prospectively through the GCTA process.

A specific example of the potential of such a registry is provided by a pilot trial to test whether ipilimumab—an antibody that inhibits immune system tolerance to

cancer cells, approved for treatment of melanoma—can act as a radiosensitizer.[76–77] There are almost certainly enough patients with melanoma already receiving radiotherapy for brain metastases, who are being treated with or without ipilimumab, that one could quickly test this hypothesis retrospectively if the data were collected and shared through a registry such as the Cancer Commons or CancerLinQ.

Technical issues and structural impediments

Panomic technologies, molecular models of cancer, targeted therapies and computational algorithms for reverse-engineering tumours are all advancing rapidly, and, except for the price of drugs, their costs are rapidly decreasing.[78] Initiatives such as CancerLinQ and Cancer Commons are creating the data bases and knowledge bases to gather and analyse the details of patient encounters. However, there are still significant technical issues, as well as structural impediments, to be overcome before patients can reap the benefits of Precision Oncology 3.0.

Technical issues

All complex biomedical technologies face the problem of potentially unreliable data and knowledge processed by sophisticated and difficult-to-validate algorithms. In the long run, the acceptance of these algorithms will depend on experimental validation. Limited formal trials of such algorithms have been conducted,[79] and we can expect to see an acceleration in experimental validation of Precision Oncology 3.0 algorithms and processes. For the foreseeable future, however, physicians bear ultimate responsibility for their patients, and so will have to make treatment decisions based upon their best judgement, potentially advised by systems such as Watson.

There are a number of technical ways to address concerns of algorithmic, data, and knowledge validity, short of mounting large scale randomized controlled trials. In the case of algorithmic validity it is common practice to use the so-called open source engineering (Box 1), wherein the entire scientific or engineering community conducts the equivalent of continuous peer review and improvement. This model has proved practical in the extremely difficult and sensitive domain of software security and privacy, including the algorithms used to secure medical software and data. In the security domain, open source algorithms are the norm (at least outside of military and espionage applications) and, indeed, are essentially demanded by the community.[80] Similarly, many journals that deal with computational analyses of biological information require that the underlying algorithms are openly revealed. The idea, although somewhat controversial,[81] is that the community as a whole can analyse and improve open algorithms, and that everyone will thereby be using, or at least will have available to hand, the most highly validated algorithms all the time.

The quality control of data and knowledge can be managed through careful accounting of the statistical certainty of the underlying evidence on which they are based. For example, results from clinical trials may be considered to be more robust or have a higher level of certainty than those from case reports, and so on all the way down to anecdotes, which presumably have very low certainty. When a treatment hypothesis is computed using this data and knowledge, the levels of certainty of the underlying data and knowledge are carried forward and accumulated in the treatment hypothesis.[82,83] As a result, hypotheses based on strong evidence will accumulate greater support, and will rank higher than those based on weaker evidence. Even subjective judgements can be incorporated into the results through this mechanism.

Structural impediments

Precision Oncology 3.0 faces significant structural impediments that include incentivizing patients and organizations to share information and materials, persuading payers to cover rational, off-label use of approved drugs, and obtaining approval to use safe INDs alone or in combination with other drugs. Many of these problems come down to regulatory policies (such as the Health Insurance Portability and Accountability Act [HIPAA]) and specific technologies (such as the automatic de-identification and re-identification of genomic data), and so are beyond the scope of this Review. However, the Rapid Learning paradigm can offer specific assistance regarding timely and affordable access to experimental and off-label drugs. This is critical to Precision Oncology because physicians will ultimately use drugs as 'molecular scalpels' to reverse engineer and treat tumours.[63]

Generally, cancer INDs are not available for use in combination therapies until they receive regulatory approval, a process that can take years, and might never happen if they fail as monotherapies.[84] Pharmaceutical companies are hesitant to provide INDs before they have significant demonstration of efficacy (for example, in animal models) because a negative result can delay the drug's regulatory approval, or shelve it entirely.[84] Moreover, although physicians can legally prescribe approved drugs for nonapproved uses, payers are reluctant to provide reimbursement without previous evidence of efficacy, preferably through standard trials.

In the short term, two emerging clinical trial models, tailored for targeted therapies, can help Precision Oncology patients get access to experimental drugs: basket trials that test targeted therapies against a 'basket' of cancer types expressing that target,[85] and bucket trials that genetically match patients to a 'bucket' of available targeted agents.[85] An excellent example, combining elements of both, is the MATCH trial proposed by the NCI. The MATCH trial aims to enroll 1,000 patients with various advanced-stage cancers who have not responded to standard treatment, test their tumours for ≥100 actionable mutations, and then match them to a large panel of targeted therapies, including INDs not yet approved for clinical use. Genomic analysis can, of course, also be used to match patients to any open trial offering the drugs they require, and pharmaceutical manufacturers and governments have been active in creating expanded access and compassionate use programmes.[86]

In the long run, precision oncologists will need the ability to prescribe off-label and combination therapies outside of formal trials. Engaging drug developers, payers, and regulatory bodies in the rapid learning process, and aligning their incentives with those of patients and physicians, could make this possible. Specifically, pharmaceutical companies could make INDs and off-label drugs available 'on approval'. On the one hand, if the treatment is successful, payers reimburse and the drug developer gets paid. Such contingent reimbursement is a common practice among European national health plans.[87] More importantly, the drug maker gets an early proof-of-concept in humans for a potentially valuable new indication. The GCTA process might then be used to quickly replicate the result in a small cohort of similar patients. In a similar way, the NCI sometimes increases dose levels in individual patients on phase I trials with a therapeutic intent (although they usually do not use the data from the higher dose level "because one cannot rule out an effect of the lower dose on the efficacy or toxicity of a subsequent, higher dose").[88] Given more time and money, a traditional trial could be mounted. However, if the treatment is unsuccessful, the knowledge gained and propagated through the network can be invaluable to future patients, and to the drug manufacturer itself, who can then reprioritize the pipeline, and the only cost is that of the tiny amount of drug used by that particular patient.

There is also the possibility that the drug will demonstrate a harmful side effect, especially an unexpected interaction in a novel combination. Without the benefit of the Precision Oncology 3.0 detailed reverse engineering process, this could be detrimental for the manufacturing of the drug. However, molecular reverse engineering could provide a rational explanation for an unexpected response. Thus, these negative results might have positive application guidance thanks to Precision Oncology rather than leading to the shelving of the drug for any indication at all.

The Rapid Learning process will obviously contribute data to cancer research, but it can, in principle, be equally well driven by hypotheses generated in the lab. Cancer researchers routinely test combinations of approved drugs or INDs on lab-grown tumours derived from specific patients. Today, if an experimental therapy elicits a dramatic response in the lab, the investigator will typically propose a clinical trial, which can take years to plan and run. Instead, under the Rapid Learning Precision Oncology model, the researcher could directly contact the very patient from whom the specimen was obtained, or patients with similar driver networks, and immediately mount either a retrospective clinical study using a centralized database, or mount a rapid prospective clinical study through the GCTA process.

Despite these issues and obstacles, interest in Precision Oncology 3.0 remains high, as evidenced by the number of recent publications and centres exploring it. It is likely, however, that at least in the near term, it will be used in parallel with traditional methods, or when all other options have been exhausted.

Conclusions

The conventional path for new anticancer drugs from the lab to the clinic extends over years via an increasingly expensive series of trials. Large, slow, drug-centric trials of this sort are no match for the immense dimensionality of the cancer problem. The therapeutically motivated experimentation of Rapid Learning Precision Oncology is predicated on the belief that the more we understand about how cancer works in each patient, the better positioned we are to help both that patient, and all that follow. Moreover, by tightly integrating research and clinical care around and across individual patients, this paradigm has the potential to dramatically accelerate knowledge acquisition and reduce delays in getting promising treatments into the clinic. Developers can get early validation of new drugs by testing them on patients with the right mutations, who are otherwise out of options. Physicians can share and learn from the thousands of clinical experiments that take place daily, but which are rarely deeply analysed, and even more rarely reported in the literature. Scientists can use preclinical experiments on a patient's cell line or xenograft to inform that patient's treatment. Most importantly, patients can be treated in accord with the best available treatments and the world's collective knowledge on how and when to use them.

To fully realize these benefits, CancerLinQ, the various Watson collaborations, Cancer Commons, and other nascent Rapid Learning Precision Oncology efforts around the world,[89] should be integrated and coordinated through a GCTA-like process to, in effect, run a huge adaptive clinical trial, whose goal is to continuously improve patient outcomes. Ultimately everyone involved in Precision Oncology should be an active participant in the Rapid Learning process so that every patient encounter counts.

Review criteria

A systematic literature search was not performed. Rather, this Review includes a summary of the authors' work and knowledge based on reading the relevant literature, attending conferences, workshops, and other meetings, and discussion with many leading researchers (cited or acknowledged).

1. Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
2. Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
3. Blau, C. A. & Liakopoulou, E. Can we deconstruct cancer, one patient at a time? *Trends Genet.* **29**, 6–10 (2013).
4. Garraway, L. A. Genomics-driven oncology: framework for an emerging paradigm. *J. Clin. Oncol.* **31**, 1806–1814 (2013).
5. Roukos, D. H. Trastuzumab and beyond: sequencing cancer genomes and predicting molecular networks. *Pharmacogenomics J.* **11**, 81–92 (2011).
6. Vallat, L. *et al.* Reverse-engineering the genetic circuitry of a cancer cell with predicted intervention in chronic lymphocytic leukemia. *Proc. Natl Acad. Sci. USA* **110**, 459–464 (2013).
7. Centre for Integrated Diagnostics. *Massachusetts General Hospital* [online], http://www.massgeneral.org/pathology/clinical/molecular.aspx (2013).
8. UW-OncoPlex—Cancer Gene Panel. *University of Washington* [online], http://web.labmed.washington.edu/tests/genetics/UW-OncoPlex (2013).

9. *Foundation Medicine* [online], http://www.foundationmedicine.com/ (2013).
10. News. *Caris life sciences* [online], http://www.carislifesciences.com/news/caris-life-sciences-to-present-on-biomarker-expression-at-the-society-of-gynecologic-oncologists-2011-annual-meeting-on-womens-cancer/ (2013).
11. *BGI* [online], http://www.genomics.cn/en/index (2013).
12. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
13. Lefebvre, C., Rieckhof, G. & Califano, A. Reverse-engineering human regulatory networks. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **4**, 311–325 (2012).
14. Carro, M. S. *et al.* The transcriptional network for mesenchymal transformation of brain tumours. *Nature* **463**, 318–325 (2013).
15. Schreiber, S. L. *et al.* Towards patient-based cancer therapeutics. *Nat. Biotechnol.* **28**, 904–906 (2010).
16. Califano, A., Butte, A. J., Friend, S., Ideker, T. & Schadt, E. Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nat. Genet.* **44**, 841–847 (2012).
17. del Sol, A., Balling, R., Hood, L. & Galas, D. Diseases as network perturbations. *Curr. Opin. Biotechnol.* **21**, 566–571 (2010).
18. Quayle, A. P., Siddiqui, A. S. & Jones, S. J. M. Perturbation of interaction networks for application to cancer therapy. *Cancer Inform.* **5**, 45–65 (2007).
19. Macconaill, L. E. & Garraway, L. A. Clinical implications of the cancer genome. *J. Clin. Oncol.* **28**, 5219–5228 (2010).
20. Bay, S. D., Shrager, J., Pohorille, A. & Langley, P. Revising regulatory networks: from expression data to linear causal models. *J. Biomedical Informatics* **35**, 289–297 (2002).
21. Shrager, J., Langley, P. & Pohorille, A. Guiding revision of regulatory models with expression data. *Pac. Symp. Biocomput.* **7**, 486–497 (2002).
22. D'haeseleer, P., Liang, S. & Somogyi, R. Genetic network inference: From coexpression clustering to reverse engineering. *Bioinformatics* **16**, 707–726 (2000).
23. Basso, K. *et al.* Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* **37**, 382–390 (2005).
24. Vaske, C. J. *et al.* Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* **26**, i237–i245 (2010).
25. Goutsias, J. & Lee, N. H. Computational and experimental approaches for modeling gene regulatory networks. *Curr. Pharm. Des.* **13**, 1415–1436 (2007).
26. Della Gatta, G. *et al.* Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumour suppressor in T-ALL. *Nat. Med.* **18**, 436–440 (2012).
27. Iyer, G. *et al.* Genome sequencing identifies a basis for everolimus sensitivity. *Science* **338**, 221 (2012).
28. Kaiser, J. Biomedicine. Rare cancer successes spawn 'exceptional' research efforts. *Science* **340**, 263 (2013).
29. Ginsburg, G. S. & Kuderer, N. M. Comparative effectiveness research, genomics-enabled personalized medicine, and rapid learning health care: a common bond. *J. Clin. Oncol.* **30**, 4233–4242 (2012).
30. Beltran, H. *et al.* A targeted next-generation sequencing of advanced prostate cancer identifies potential therapeutic targets and disease heterogeneity. *Eur. Urol.* **63**, 920–926 (2013).
31. Meric-Bernstam, F., Farhangfar, C., Mendelsohn, J. & Mills G. B. Building a personalized medicine infrastructure at a major cancer centre. *J. Clin. Oncol.* **31**, 1849–1857 (2013).
32. Balbin, O. A. *et al.* Reconstructing targetable pathways in lung cancer by integrating diverse omics data. *Nat. Commun.* http://dx.doi:10.1038/ncomms3617 (2013).
33. Mount Sinai to Test. *Cancer Network Modeling Approach in Two Clinical Trials March 27*, *GenomeWeb* [online], http://icahn.mssm.edu/static_files/MSSM/Genomics/Clin.Seq_050913_2.pdf (2013).
34. *GeneKey* [online], http://www.genekey.com/.
35. Tailored clinical molecular test interpretation. *N-of-One* [online], http://www.n-of-one.com/.
36. Olsen, L. A., Aisner, D. & McGinnis, J. M. The Learning Healthcare System: workshop summary (Institute Of Medicine roundtable on evidence-based medicine). The National Academies Press (2007).
37. Etheredge, L. M. A rapid-learning health system. *Health Aff.* **26**, w107–w118 (2007).
38. Murphy, S. & Patlak, M. A foundation for evidence-driven practice: a rapid-learning system for cancer care: workshop summary. The National Academies Press (2009).
39. Abernethy A. P. *et al.* Rapid-learning system for cancer care. *J. Clin. Oncol.* **28**, 4268–4274 (2010).
40. Fenstermacher, D. A., Wenham, R. M., Rollison, D. E. & Dalton, W. S. Implementing personalized medicine in a cancer centre. *Cancer J.* **17**, 528–536 (2011).
41. Ren, Y. *et al.* JAK1 truncating mutations in gynecologic cancer define new role of cancer-associated protein tyrosine kinase aberrations. *Sci. Rep.* **3**, 3042 (2013).
42. Fox, E. J., Salk, J. J. & Loeb, L. A. Cancer genome sequencing: an interim analysis. *Cancer Res.* **69**, 4948–4950 (2009).
43. Welch, J. S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
44. Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* **150**, 251–263 (2012).
45. Weinberg, R. A. *The Biology of Cancer*. (Garland Science, New York, 2007).
46. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nat. Med.* **10**, 789–799 (2004).
47. Pathways to Cancer. *Inside Teacher Cancer Centre* [online], http://teachercentre.insidecancer.org/browse/Pathways%20to%20Cancer/ (2013).
48. Lamb, J. *et al.* The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
49. Ellis, M. J. *et al.* Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* **486**, 353–360 (2012).
50. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups *Nature* **486**, 346–352 (2012).
51. Lundberg, G. D. Medical Publishing for an N of One. *The Scientist* **25**, 31 (2011).
52. National Research Council. Committee on a framework for development a new taxonomy of disease. Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease. The National Academies Press (2011).
53. *PLoS Currents* [online], http://currents.plos.org/ (2013).
54. Ram, K. Git can facilitate greater reproducibility and increased transparency in science. *Source Code Biol. Med.* **8**, 7 (2013).
55. Gallo, P. *et al.* Adaptive designs in clinical drug development–an executive summary of the PhRMA working group. *J. Biopharm. Stat.* **16**, 275–283 (2006).
56. Fiore, L. D. *et al.* A point-of-care clinical trial comparing insulin administered using a sliding scale versus a weight-based regimen. *Clin. Trials* **8**, 183–195 (2011).
57. Kim, E. S. *et al.* The BATTLE Trial: Personalizing therapy for lung cancer. *Cancer Discov.* **1**, 44–53 (2011).
58. Berry, D. A. Adaptive clinical trials in oncology. *Nat. Rev. Clin. Oncol.* **9**, 199–207 (2012).
59. D'Avolio, L. *et al.* Implementation of the Department of Veterans Affairs' first point-of-care clinical trial. *J. Am. Med. Inform. Assoc.* **19**, e170–e176 (2012).
60. Scher, H. I., Nasso, S. F., Rubin, E. H. & Simon, R. Adaptive clinical trial designs for simultaneous testing of matched diagnostics and therapeutics. *Clin. Cancer Res.* **17**, 6634–6640 (2011).
61. Shrager, J. Theoretical Issues for Global Cumulative Treatment Analysis (GCTA). *Cornell University Library* [online], http://arxiv.org/abs/1308.1066 (2013).
62. Vickers, A. J. & Scardino, P. T. The clinically-integrated randomized trial: proposed novel method for conducting large trials at low cost. *Trials* **10**, 14 (2009).
63. Huber, P. The digital future of molecular medicine: rethinking FDA regulation. Project FDA report no. 6. *The Manhattan Institute* [online] http://www.manhattan-institute.org/pdf/fda_06.pdf (2013).
64. Shrager, J. & Tenenbaum, J. M. Cancer Commons: Biomedicine in the internet age. In *Collaborative Computational Technologies for Biomedical Research* (ed. Elkin, S.) 161–177 (John Wiley & Sons, 2011).
65. Tenenbaum, J. M. & Shrager, J. Cancer: A computational disease that AI can cure. *AI Magazine* **32**, 14–26 (2011).
66. Sledge, G. W. *et al.* ASCO's Approach to a Learning Health *Care* System in Oncology. *J. Oncol. Pract.* **9**, 145–148 (2013).
67. ASCO's CancerLinQ: Building a Transformation in Cancer Care. *ASCO* [online], http://www.ascopost.com/issues/july-1-2012/asco%E2%80%99s-cancerlinq-building-a-transformation-in-cancer-care.aspx (2013).
68. ASCO Completes Prototype for CancerLinQ™, Marking First Demonstration of a "Learning Health System" to Transform Cancer Care. *ASCO* [online], http://www.asco.org/press-centre/asco-completes-prototype-cancerlinq%E2%84%A2-marking-first-demonstration-%E2%80%9Clearning-health (2013).
69. Malin, J. L. Envisioning Watson as a rapid-learning system for oncology. *J. Oncol. Pract.* **9**, 155–157 (2013).
70. IBM Watson helps fight cancer with evidence-based diagnosis and treatment suggestions. *Memorial Sloan-Kettering Cancer Centre* [online], http://www-03.ibm.com/innovation/us/watson/pdf/MSK_Case_Study_IMC14794.pdf (2013).
71. Vidwans, S. *et al.* A melanoma molecular disease model. *PLoS ONE* **6**, e18257 (2011).
72. West, L. *et al.* A novel classification of lung cancer into molecular subtypes. *PLoS ONE* **7**, e31906 (2012).
73. Participate in rapid learning communities. *Cancer Commons* [online], http://www.cancercommons.org/researchers-clinicians/ (2013).
74. Donate Your Data. *Cancer Commons* [online], http://www.cancercommons.org/donate-your-data/ (2013).

75. Grove, A. Rethinking clinical trials. *Science* **23**, 1679 (2011).
76. US National Library of Medicine. *ClinicalTrials.gov* [online], http://clinicaltrials.gov/ct2/show/NCT01449279 (2013).
77. Wolchok, J. D. Emerging therapies in melanoma. *Clin. Adv. Haematol. Oncol.* **10**, 742–743 (2012).
78. DNA Sequencing Costs. *National Human Genome Research Institute* [online], http://www.genome.gov/sequencingcosts/ (2013).
79. Monks, N. R. *et al.* A multi-site feasibility study for personalized medicine in canines with Osteosarcoma. *J. Transl. Med*. **11**, 158 (2013).
80. Hoepman, J. H. & Jacobs, B. Increased security through open source. *Commun. ACM* **50**, 79–83 (2007).
81. Anderson, R. J. Open and closed source systems are equivalent (that is, in an ideal world). In Perspectives on free and open source software. 127–142 (The MIT Press, Cambridge, MA, 2005).
82. Shrager, J., Billman, D., Convertino, G., Massar, J. P. & Pirolli, P. Soccer science and the Bayes community: exploring the cognitive implications of modern scientific communication. *Top. Cog. Sci.* **2**, 53–72 (2009).
83. Domingos, P. & Webb, W. A. A tractable first-order probabilistic logic [abstract]. Proceedings of the 26th AAAI Conference on Artificial Intelligence. Toronto, Canada (2012).
84. Humphrey, R. W. *et al.* Opportunities and challenges in the development of experimental drug combinations for cancer. *J. Natl Cancer Inst.* **103**, 1222–1226 (2011).
85. Willyard, C. 'Basket studies' will hold intricate data for cancer drug approvals. *Nat. Med*. **19**, 655 (2013).
86. Lerose, R., Musto, P., Aieta, M., Papa, C. & Tartarone, A. Off-label use of anti-cancer drugs between clinical practice and research: the Italian experience. *Eur. J. Clin. Pharmacol.* **68**, 505–512 (2012).
87. McCabe, C. *et al.* Market and patient access to new oncology products in Europe: a current, multidisciplinary perspective. *Ann. Oncol*. **20**, 403–412 (2009).
88. Doroshow, J. H. Personal communication to Jeff Shrager. 2013-08-14.
89. Cancer Research UK. *Stratified Medicine Programme* [online], http://www.cancerresearchuk.org/science/research/how-we-deliver-our-research/others/by-programme/stratified-medicine-programme/ (2013).

Acknowledgements

We especially thank C. A. Blau for his significant contributions to both the concepts and content of this paper. Others who provided significant feedback include A. Califano, O. Capote, J. Doroshow, G. S. Ginsberg, S. Greene, P. Howard, P. Huber, R. Lehrer, L. Marton, G. B. Mills, K. Polacek, A. Schiffman, E. Shtivelman, and several anonymous reviewers.

Author contributions

Both authors researched data for article, made a substantial contribution to discussion of the content, and wrote and edited the manuscript before submission.