# The Targeted Therapy Analyzer (TTA): A literature-based tool for the targeted therapy of melanoma

# Jeff Shrager, Ph.D, Randy Gobbel, Ph.D, and Simon Mocellin, M.D.

# Mocellin, et al. (2010) describe the “The Targeted Therapy Database” (TTD), a publicly-available repository that provides investigators with a searchable and computation-compatible collection of the scientific evidence regarding the targeted therapy of melanoma. They also specify, and implement by example in Excel, two algorithms, one that “identifies prevalent therapeutic hypotheses”, and a second that ranks potential therapeutic choices based upon these hypotheses and whether a particular tissue’s molecular profile is concordant with the assays that index those hypotheses. Together these are called by Mocellin, et al. (2010) “Targeted Therapy Analysis”, or TTA. The present letter describes a web-based implementation of the TTA, which makes them quite a bit easier to use and more widely accessible. In the next sections we briefly describe the TTA – readers are encouraged to try it out for themselves in order to get a better sense of how it works and is used. Following this, we describe limitations of the current TTA, and possible approaches to surmount these.

### http://www.imedicalapps.com/2011/02/collabrx-provides-web-applications-for-personalized-cancer-therapy-a-vision-of-the-future-2/

# “[CollabRx provides web apps for selecting targeted therapies: the future of personalized cancer treatment?](http://www.imedicalapps.com/2011/02/collabrx-provides-web-applications-for-personalized-cancer-therapy-a-vision-of-the-future-2/)

by [Felasfa Wodajo, MD](http://www.imedicalapps.com/author/felasfa/)

on 02/17/11

ShareThis

The rapid proliferation of knowledge about the molecular underpinnings of different cancers has brought hope for a new age of “targeted” therapies. These drugs are designed to find and destroy cells with aberrant biochemical pathways, while bypassing the normal body tissues. Immense hopes rest on them.

However, if managing the terrible toxicities of traditional “cytotoxic” chemotherapy is the special skill of today’s oncologist, the future holds new complexities for the chemotherapist. This will be to concoct the best match between the pattern of genetic mutations in a particular tumor with the available targeted therapies. Today, the number of clinically relevant mutations that are known and the number of FDA approved medications is not large. However, the rapid pace of discovery will likely render this early period short. This is already becoming the case with melanoma, where up to four molecular therapies are in various stages of FDA approval, and at least three different mutations are targets. For melanoma, CollabRx now provides a simple a web application to make this match at [http://therapy.collabrx.com](http://therapy.collabrx.com" \t "_blank). The user enters the types of mutations found in the tumor plus some other information, and the application generates a suggested therapy. Again, at this time, the number of potential combinations is not large. However, the model is novel and suggests a different future in a few respects. First, the treatment recommendation is specific to that individual. This is not population based data meant to serve as a clinical guideline. In fact, by also creating a [non-profit wiki “cancer commons”](http://mmdm.cancercommons.org/smw/index.php/A_Melanoma_Molecular_Disease_Model" \t "_blank), the company hopes to grow the edifice of actionable information faster than the current method of single peer-reviewed journal article at a time.

Second, the company designed the crowd-sourced “molecular disease model (MDM)” as a structured or “semantic” document. The hope is that MDMs for other diseases will be created in the future and, using what they term “computational knowledge representation”, they can build similar applications for other diseases. Last and fundamental, the same application is available to both physicians and patients. If great hopes rest on the ability of information sciences to support decision making by physicians, there is no reason the same cannot hold true for patients. Opening avenues of inquiry to the vast number of highly motivated patients can only accelerate the pace of advance. It is also interesting that the company was founded by Dr. Jay (Marty) Tenenbaum who was living in the shadow of his own battle of advanced melanoma. He was Stanford PhD who was ahead of his time in anticipating the potential for the internet for commerce. His company Commerce One was hugely successful and he used some of his fortune to start CollabRx. Some readers may remember that he was also instrumental in the creation of the [Health Apps Accelerator](http://www.imedicalapps.com/2011/01/interview-with-founding-member-of-health-2-0-accelerator-erick-von-schweber-ceo-of-surveyor-health/), a framework for connecting clinical web “apps”. While the CollabRx on-line therapy tool is not strictly a mobile medical app at this time, according to Jeffrey Shrager, CTO, it is likely that “[CollabRx] will soon be distributing them through mobile and/or embedded model”. By bringing the benefits directly to individual, this tool echoes of the way mobile devices have empowered physicians and patients alike. Beyond that, it suggests how information technologies and clinical medicine may be combined in the future to create more powerful tools.”

==========

http://www.medpagetoday.com/Columns/24731

Hello and welcome. I'm Dr. George Lundberg and this is At Large at MedPage Today.

What with the iPads and generations of iPhones and their competitors, most people recognize that "apps" no longer only stands for apple fruit, the Apple company, or even surgery for appendicitis.

"Apps" stands for applications, which in medicine can be decision support tools distributed for use on peripheral devices, not a traditional central website.

I expect a proliferation of clinically useful "apps" to guide physicians and patients together to the best diagnostic and therapeutic decisions.

One such app is now available for testing and use in the new world of molecular oncology. My gross and microscopic skills in anatomic pathology are no longer sufficient for the proper diagnosis of many cancers.

Approximately 70,000 Americans are diagnosed with melanoma each year and about 7,000 die, after proceeding beyond "standard of care."

Newer molecular/genomic tests offer hope for some patients with metastatic melanoma, but only if their cancers are tested to help guide the patient into the best clinical trials or the most promising targeted therapy.

A company called CollabRx has recently made available a clinical diagnostic app, called the Targeted Therapy Finder -- Melanoma. It is free of charge at [http://therapy.collabrx.com](http://therapy.collabrx.com/" \t "_blank).

This app is based on the evolving science of the Melanoma Molecular Disease Model, developed under the leadership of Cancer Commons-Melanoma co-chief editors David Fisher and Keith Flaherty of Harvard Medical School. As you may know, I am also the editor-in-chief of Cancer Commons at [www.cancercommons.org](http://www.cancercommons.org/" \t "_blank).

This interactive app plays on a semantic wiki. Try it.

Start with the melanoma types and locations, simple staging, status of molecular tests, and follow them to the gene/mutation tables, pathways, available diagnostic tests, and labs where they are performed, best clinical trials, and potential treatments.

This app launches a new future for cancer education and treatment by blending molecular oncology with a semantic Web wiki. We hope you find it helpful to use and teach.

Let us know your criticisms so this "work in progress" can be improved.

Cancer kills around 600,000 Americans every year, and we have not made much progress at lowering that mortality over the last 40 years. We hope that blending genomic science and advanced information technology can begin to improve those outcomes.

=================

http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0011965

The Targeted Therapy Database is the first publicly available repository that provides investigators with a searchable and computation-compatible collection of the scientific evidence regarding the targeted therapy of melanoma. Users can query the database to easily obtain standardized information about the molecular determinants of sensitivity or resistance of melanoma to a given treatment, the compounds that can synergize with a given treatment, as well as the molecular determinants of toxicity of a given treatment.

This information can be utilized to quickly ascertain the most studied as well as the emerging therapeutic strategies, along with the models where they have been tested and the results yielded so far.

Using the above presented model based on the evidence score, these data can also be exploited As above explained, although our model cannot quantify the therapeutic benefit of a given targeted therapy, it can be used to discern trends in the available evidence, pinpointing the most promising approaches based on the amount of literature (rated according to the scoring method described above) in favor of each therapeutic hypothesis.

Finally, this archive - along with the algorithm we have proposed - can be utilized to match the patient's molecular profile with the available literature and thus to hypothesize patient-specific drug sensitivity toxicity or synergism based on the scientific evidence supporting each type of relationship for each of the molecules investigated.

We chose melanoma because this tumor paradigmatically represents the urgency of providing patients with better treatments: in fact, no current drug regimen significantly impacts on the clinical course of this disease in the metastatic setting. Under these unfortunate circumstances, any therapeutic choice based on the available evidence (even without clinical proof of efficacy of such a strategy) would appear more rational than offering patients no options at all. However, since the drug ranking system described above is based on a theoretical model, it should only be used to generate hypotheses, not to make clinical decisions. In other words, at the moment the findings obtained with our model should only be used a posteriori (after the patients has been treated with a regimen chosen independently of the model results) in order to determine the actual performance of the model itself. Only this validation of the model on the clinical ground will enable us to verify whether our theoretical computations are accurate enough to be clinically valuable, and thus to propose the implementation of the model in the routine setting for choosing the therapeutic regimen most likely to benefit individual patients.

Despite its intrinsic limitations (e.g., the score is arbitrary, the literature coverage is incomplete and thus many hypothesis are based on few or even single original articles), this model is - to the best of our knowledge - the first attempt to directly apply the enormous amount of data accumulated by the scientific community in the field of personalized medicine. This translational approach has the undeniable advantage of making the most of the scientific production by using it comprehensively, without wasting any evidence. This can be envisaged as an effort to deal with the general problem that the biomedical community produces more data than those utilized for clinical purposes. The actual impossibility of testing each preclinical hypothesis in the clinical setting represents undoubtedly a waste of potentially useful information: this “abandoned” information could be “rescued” by taking it into consideration through the model we propose for the evidence-based design of further research, both preclinical and clinical. Should the clinical validation of this drug ranking system demonstrate that it is reliable, the TTD could be utilized as a template to develop similar repositories dedicated to any tumor and more generally to any disease.

On the other hand, it should be clearly noted that scoring the hypotheses reported in the literature as we propose to do here cannot replace the standard rules of research, including clinical phases of treatment evaluation and formal meta-analysis of therapeutic interventions. The model we presented can only speed up the identification of the most promising hypotheses of targeted therapy by making an unprecedented comprehensive use of the available evidence based on two principles: 1) any information is potentially useful, independently of the experimental model that has generated it, provided that different “weights” are assigned to different models in order to reflect the difference in reliability; 2) disease's outcomes virtually always depend upon molecular combinations, which calls for the simultaneous use of information about all the molecules so far investigated, which should maximize the likelihood of successfully drive targeted therapies.

As the available and eligible data are added to the TTD, we will be able to make predictions more and more reliable because they will be based on more information. In particular this will minimize the risk of publication bias because some positive/significant molecular associations published in the first place will be “balanced” by negative/non significant findings. We note that - in analogy to standard meta-analysis - the greater the number of studies considered the smaller the variance of the overall effect; in our case, the smaller the sampling error the more accurate the prediction. Furthermore, the growing information will enable investigators to make setting specific predictions thanks to the flexibility of the TTD: in fact, its format allows to insert more columns (e.g., a new one could be dedicated to distinguish data obtained in the primary tumor or metastatic setting) at any time. Then our model can still be applied as above described because the user can simply sort the database by the new column (e.g., primary vs. metastatic) and use only the relevant information (e.g., data from primary or metastatic setting) based on the clinical question to be addressed.

Finally, we would like to underscore that this kind of project can succeed only if the scientific community participates in the effort of improving the model we have proposed. This can be realized in several ways, such as: A) by giving notice of relevant articles not yet included in the TTD, which will maximize the literature coverage of the database and thus will ultimately increase the reliability of the analyses performed; B) by proposing new algorithms improving the exploitation of the information contained in the database; C) most importantly, by testing the hypotheses generated by the TTD analyses both in the preclinical and clinical setting.

Overall, putting together the pieces of a “disease puzzle” is becoming increasingly difficult due to the continuous and growing flow of information that no single mind can keep up with: we therefore propose the TTD (and the associated model for drug ranking) as a tool for the synopsis and synthesis of the scientific hypotheses with the aim of favoring the rational design of both preclinical and clinical research.

The commitment of the MMMP Team (the core of basic researchers and clinical investigators taking care of the scientific content of the MMMP website) is not only to keep the TTD regularly updated but also to carefully take into consideration suggestions, criticisms and contributions from the scientific community.

We strongly believe that the bidirectional exchange of information (from the database to the user and vice versa) represents the most efficient way of gathering and exploiting scientific data on a specific disease: in fact, if every researcher spent just a small amount of time to share his/her knowledge to keep up-to-date the TTD or any other similar project, the pace of discovery of more effective anticancer strategies would be greatly increased.

================

Theobald, M., Shah, N., Shrager, J. (2009). Extraction of Conditional Probabilities of the Relationships between Drugs, Diseases, and Genes from PubMed guided by relationships in PharmGKB. AMIA Summit on Translational Bioinformatics, San Francisco, CA.

**Abstract**

Guided by curated associations between *genes*, *treatments* (i.e., drugs), and *diseases* in pharmGKB, we constructed n-way Bayesian networks based on conditional probability tables (cpt’s) extracted from co-occurrence statistics over the entire Pubmed corpus, producing a broad-coverage analysis of the relationships between these biological entities. The networks suggest hypotheses regarding drug mechanisms, treatment biomarkers, and/or potential markers of genetic disease. The cpt’s enable Trio, an inferential database, to query indirect (inferred) relationships via an SQL-like query language.

We seek to extract all-way co-occurrence-based Bayesian networks among treatments (primarily drugs for this study), diseases, and genes. These can be estimated from subsets of conditional and non-conditional probabilities which are in turn derived from raw co-occurrence counts of drug/disease/gene entities in domain-specific corpora such as Pubmed. For non-conditional statistics, such a co-occurrence probability would simply be the number of documents (or abstracts) that mention these items together, divided by the total number of documents contained in the corpus. The desired conditional probabilities are: p(drug|gene), p(drug|disease), p(drug|gene,disease), etc. One can easily see how to compute such conditional probabilities over an appropriately annotated Pubmed database, simply by counting the single and combinational co-occurrences of all of these entities, and performing the obvious calculation, i.e., p(drug A|gene B, disease C) = (# distinct abstracts containing A and B and C)/(# distinct abstracts containing B and C). Notice that more general relationships are conceivable, i.e., considering many-to-many relationships between drugs, diseases, and genes. In the present experiment we limit our Bayesian network to a maximum of six conditional variables and a single target variable, thus extracting up to 26 conditional probabilities per net. This keeps the combinatorial complexity, and hence the number of co-occurrence queries issued against our underlying Pubmed corpus, reasonable.

Guided by the relations in pharmGKB 1 , we combined information from a tagged Pubmed corpus created by processing all Medline abstracts2 using the Mgrep tool (University of Michigan). Mgrep uses all of the alternative strings for UMLS concepts3 and identifies their occurrence in the abstract using a radix tree based method that allows for very fast processing without sacrificing precision [5]. In our experience, this method has an average precision of about 85% for diseases [6]. (We have not evaluated precision for other entities.)

The tagged corpus used in the present experiment contains ~19 million articles and ~200 concepts assigned to each article resulting in ~3 billion unique Pubmed ID-to-UMLS Concept Unique Identifier (CUI) assignments. We combined these data with the highly reliable gene2pubmed database (ftp.ncbi.nih.gov/ gene/DATA). Using only the relationships marked as “related” or “positively related”, we extracted 1,730 disease/drug/gene relationships with up to 6 conditional variables, and extracted their respective conditional probability tables using co-occurrence statistics over the ~3 billion distinct Pubmed ID/CUI pairs, resulting in 19,092 conditional probabilities (again, compare with ~222,000 for the full-joint distribution). Although this is clearly still an offline process, requiring several days, once extracted, these tables serve as input for our Bayesian nets and allow for an efficient execution of arbitrary inferential queries; *any conditional probability of variables/entities expressed in a pharmGKB relationship can be directly computed from these*.

The probabilities that we derive reflect only co- occurrence in the literature, and *not*, for example, recommendations, so one must be cautious in interpreting these results. What, then, are they telling us, and is what they are telling us useful? Because the literature is historical, these results are not telling us what to try, but *what has been tried*, or, possibly, *what has been suggested* (if not actually tried).

==================