

# Accelerating Cancer Research

Enabling Real-Time Discovery  
in Big Data

## Urika Advantage:

In order to rapidly validate scientific hypotheses in real time and discover new connections within their existing data, the ISB team needed a powerful solution that enabled data discovery at scale.

## The Challenge

Cancer researchers have a wealth of data available to them regarding the molecular and clinical characteristics of the many forms of cancers and the use of therapeutic drugs to treat disease, including both proprietary research from their own labs as well as publicly available data such as The Cancer Genome Atlas and other collaborative scientific and public data sources.

The hypothesis is that Big Data could be used to identify potential new drug treatments from data already available to researchers by analyzing gene-drug relationships within cancer and other therapeutic areas without performing “wet” lab work first.

However, traditional analytics tools and techniques to test these hypotheses often take several weeks to months to execute. They are time consuming because data scientists must assemble all of the necessary data into a new data model to determine whether the researcher’s hypothesis is accurate. Because of the extensive amount of time between question and answer, the experiment’s results may be irrelevant by the time they are finally delivered.

The researchers at the Institute for Systems Biology wanted to determine if they could significantly reduce the lengthy time period between posing a hypothesis regarding possible cancer treatments and receiving answers. They wanted a way to get to yes-or-no quickly in order to prioritize drug re-purposing opportunities; this would then accelerate the discovery of new cancer treatments that can be moved through the drug development and approval process quickly, thus making a major difference to cancer patients.

The impact of using a more powerful analytics solution was immediate—and dramatic: In the amount of time it previously took to validate a single hypothesis, the team could now validate 1,000.

## The Solution

The ISB team worked with YarcData to develop an innovative, real-time approach to cancer research discovery using the Urika graph analytics appliance. Using Urika, the team was able to assemble all of their data into a single graph in the appliance's vast shared memory—eliminating the need to partition the data or create time-consuming and complex data models prior to posing a hypothesis. This solution is scalable, which allows the data set to expand over time without losing performance or data integrity.

The ISB team identified new cancer therapy candidates by exploring correlations between frequently mutated genes from tumor samples in order to identify existing gene-drug associations that could be possible drug candidates. In addition to discovering promising new therapies, they also sought to rapidly eliminate the large number of drugs that would not be useful in delivering the desired result.

In order to deliver results quickly, the researchers needed a way to discover unknown relationships within the data that the current data management strategy couldn't deliver. Urika enabled ISB's researchers to look at the data differently than query-based relational database systems where the data determined what investigative questions could be asked. This resulted in a clear visualization of the connections and associations within the data to help identify promising candidates for new therapies.

The graph analytics approach enabled the research team to identify thousands of drug repurposing opportunities that warrant further investigation. For example, this methodology identified a drug that is currently used to treat HIV called Nelfinavir, which showed selectivity in a separate research study for HER2-breast cancer. The ISB team came to the same conclusion about Nelfinavir in a fraction of the time, with no need for hands-on "wet lab" work to test the hypothesis—validating the accuracy and efficacy of the "big data" approach for identifying drug treatment solutions.

## The Urika Advantage

Urika, with its large global shared memory, RDF/SPARQL interface and proprietary Threadstorm multithreaded graph processors, allowed the team to rapidly integrate ISB's proprietary data with publicly available data, enabling the researchers to identify new relationships in the data without any upfront modeling. No advance knowledge of the relationships within the data is required to identify non-obvious patterns, facilitating true data discovery.

When compared to traditional database strategies and investigative laboratory experiments, the ISB researchers were able to significantly reduce the time to discovery, saving months or years of research with a higher probability of success.



**About the Urika™** The YarcData Urika big data appliance for graph analytics helps enterprises gain business insight by discovering relationships in big data. Urika complements an existing data warehouse or Hadoop cluster by offloading graph workloads and interoperating within the existing analytics workflows. Subscription pricing or on-premise deployment of the appliance eases Urika adoption into existing IT environments.

**About YarcData** YarcData, a Cray company, delivers business-focused real-time graph analytics. Adopters include the Institute of Systems Biology, the Mayo Clinic, Noblis, Sandia National Labs, as well as multiple deployments in the US government. YarcData is based in the San Francisco bay area and more information is available at [www.yarcdata.com](http://www.yarcdata.com).

©2013 Cray Inc. All rights reserved. Specifications subject to change without notice. Cray is a registered trademark, YarcData and Urika are trademarks of Cray Inc. All other trademarks mentioned herein are the properties of their respective owners.