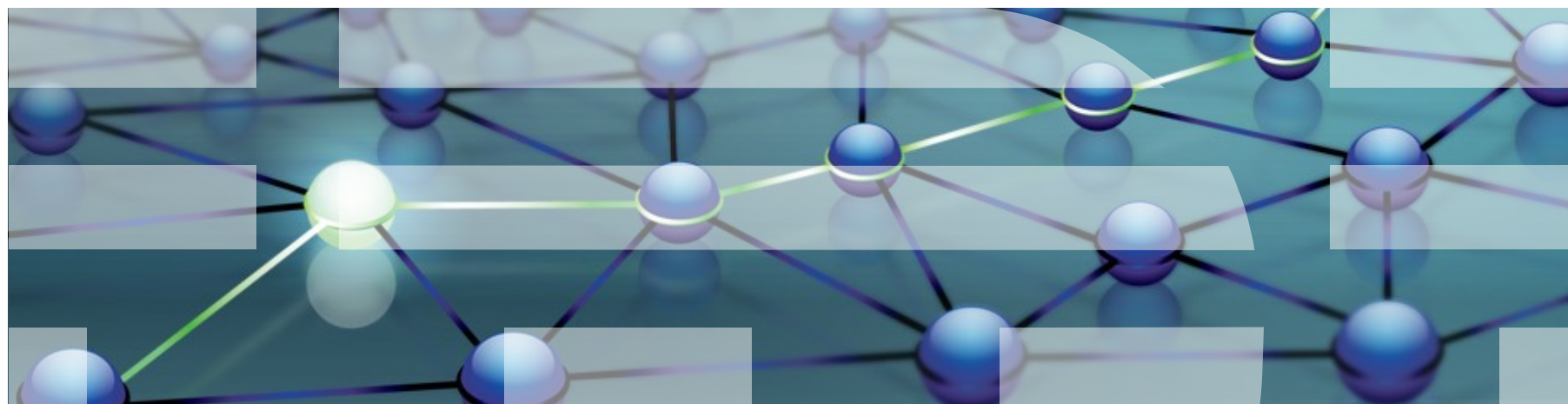# Watson Discovery Advisor

## Predicting Future Scientific Discoveries Based on Analysis of Past Literature

Scott Spangler, Principal Data Scientist

# AUTOMATED HYPOTHESIS GENERATION BASED ON MINING SCIENTIFIC LITERATURE

Scott Spangler, Angela D. Wilkins, Benjamin J. Bachman, Meena Nagarajan, Tajhal Dayaram, Peter Haas, Sam Regenbogen, Curtis R. Pickering, Austin Comer, Jeffrey N. Myers, Ioana Stanoi, Linda Kato, Ana Lelescu, Jacques J. Labrie, Neha Parikh, Andreas Martin Lisewski, Lawrence Donehower, Ying Chen, Olivier Lichtarge

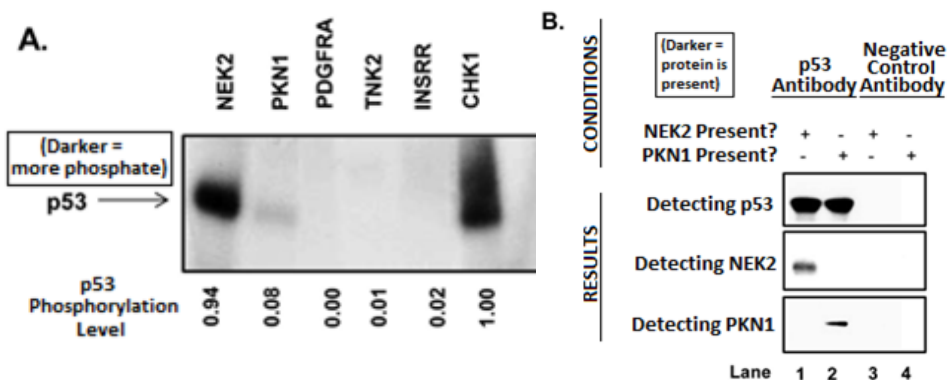## Product of collaboration between IBM and Texas Medical Center (BCM, MDA, RiceU)
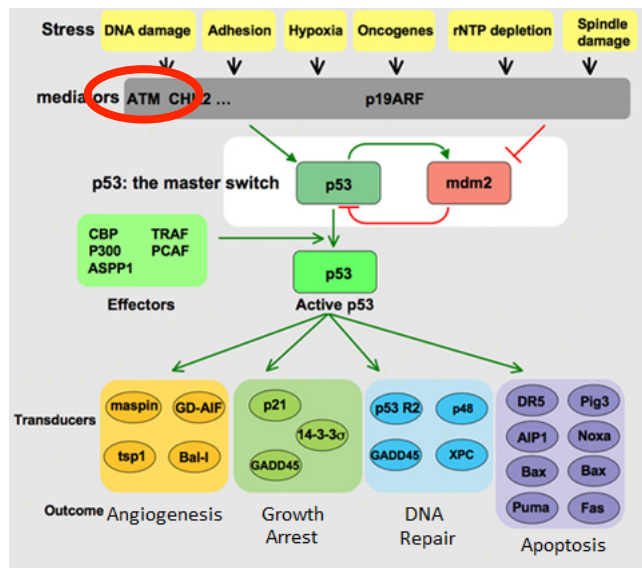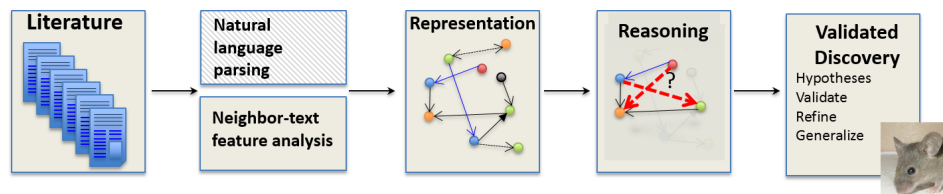
Aim: Find controlling proteins in cancer pathways

GREEN – p53 Kinases
RED/ORANGE/YELLOW – Predicted Targets
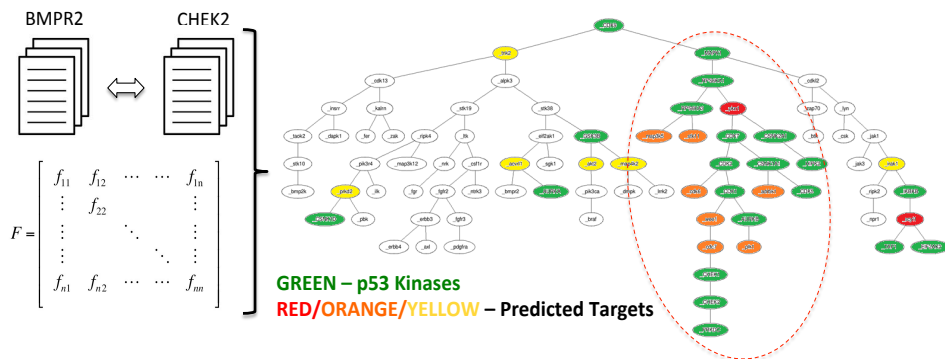
Given kinases with known function, we can make novel predictions based on connections

## Experimental results confirm functional predictions

- **Normally found at a rate of ~ 1/year**
- **Potential cancer treatment targets**
- Proof-of-principle for grander scale

High Level Process for Accelerated Discovery

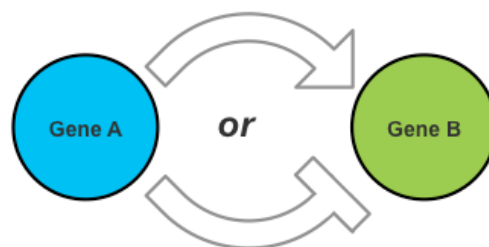## Function

Step 4: Prediction
Put all entities and relationships together in context to form a picture of what's going on and predict downstream effects

Known Pathways   **Predicted Effects**
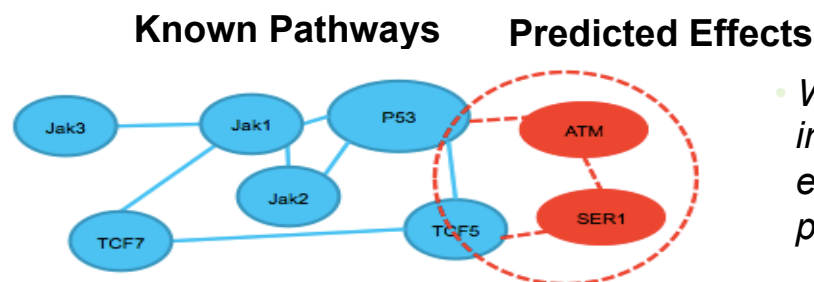
- *What are the implications of protein effects on disease pathways*

Step 3: Relationships
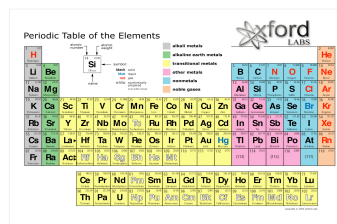How do entities influence and affect one another in specific situations

Look for protein interactions for a given set of proteins to model potential biological pathways

Form

Step 2: Organization
Gain deep understanding of all domain-relevant concepts and relationships

Ontologies
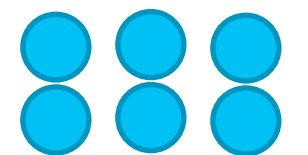**(e.g. organism, cell, protein, amino acid)**

Step 1: Exploring for Entities
Explore rich and diverse domain content for patterns

Unstructured

Entity Recognition & Normalization
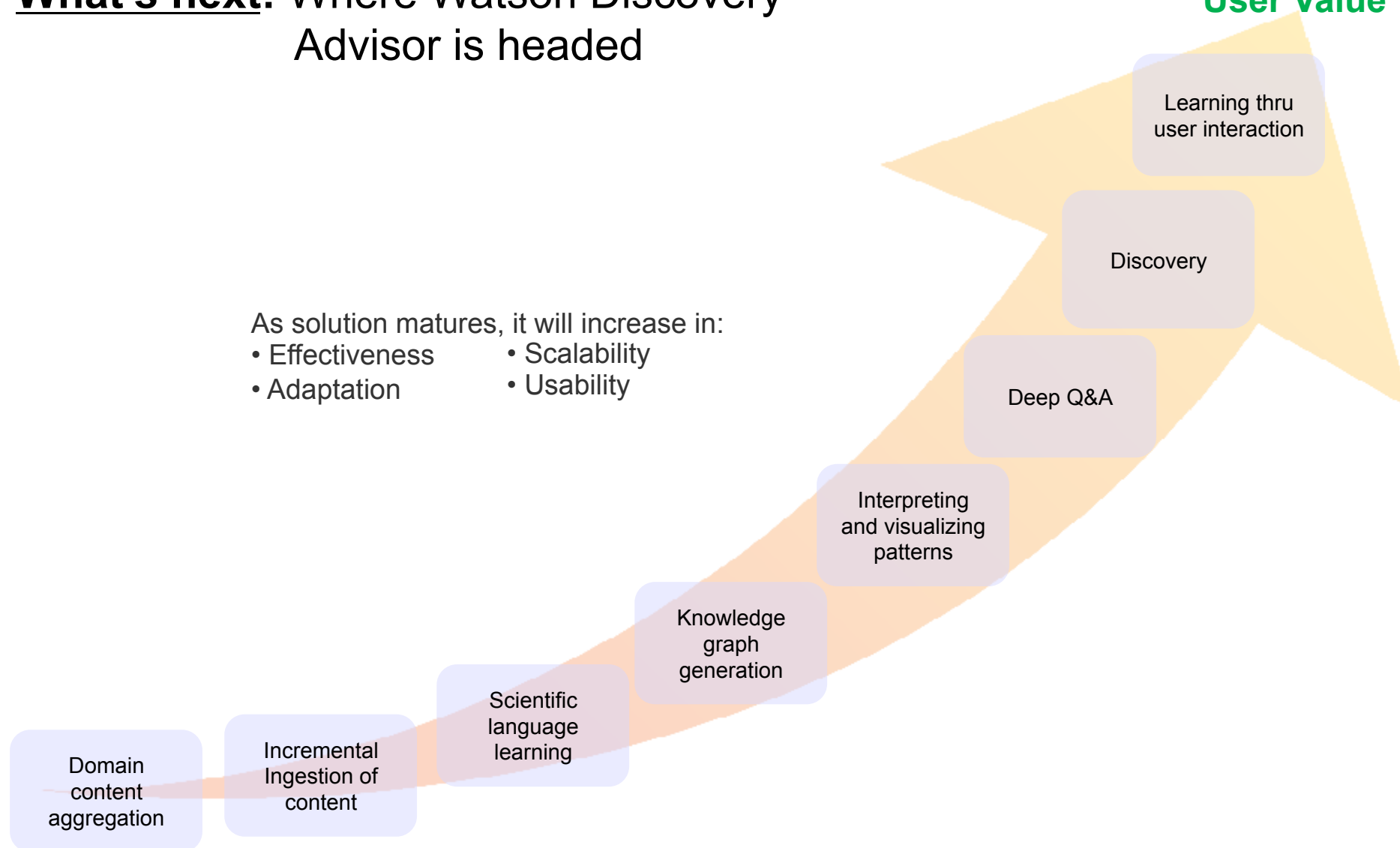
Domain Entities

IBM WATSON

What's next: Where Watson Discovery Advisor is headed

User Value

As solution matures, it will increase in:
- Effectiveness
- Adaptation
- Scalability
- Usability

Learning thru user interaction

Discovery

Deep Q&A

Interpreting and visualizing patterns

Knowledge graph generation

Scientific language learning

Incremental Ingestion of content

Domain content aggregation

Content may change as we engage additional customers