QUANTORI

Unveiling the Multimodal Frontier: AI-Driven Hypothesis Generation in Drug Discover

Steven E. Labkoff, MD, FACP, FACMI, FAMIA

Global Head, Healthcare and Clinical Informatics

Outline

Introduction to multimodal data in drug discovery

AI's role in hypothesis generation

Case studies and current work in the field

Challenges in AI Drug Design

Q&A

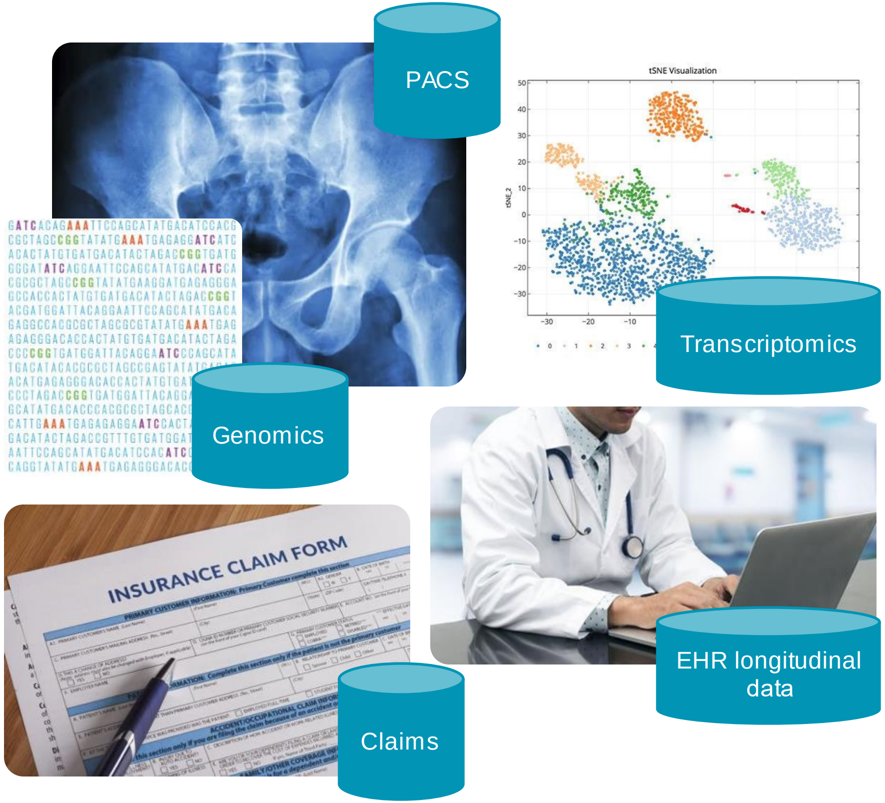
AI in Drug Discovery: A Paradigm Shift

QUANTORI

Traditional Drug Discovery	AI Driven Drug Discovery
Linear and Sequential: Processes often follow a linear path from target identification to lead optimization.	**Parallel and Agile:** AI can concurrently analyze multiple stages of drug development.
Hypothesis-Driven: Relies on preconceived hypotheses based on prior knowledge and experience.	**Data-Driven Hypotheses:** Generates hypotheses from vast datasets, identifying patterns beyond human recognition.
Manual Data Analysis: Data analysis is often manual, time-consuming, and subject to human error.	**Automated Data Analysis:** Leverages machine learning to process and analyze large datasets quickly.
Limited Data Integration: Struggles to integrate diverse data types due to methodological silos.	**Multimodal Data Utilization:** Capable of integrating heterogeneous data types (e.g., genomics, proteomics) for a holistic view.
Slower Iteration: Each phase of drug development can take years, with slow feedback loops.	**Rapid Iteration Cycles:** AI algorithms can quickly learn from data, enabling faster optimization.
High Attrition Rates: Many compounds fail late in development due to lack of efficacy or unforeseen toxicity.	**Predictive Analytics:** Uses predictive models to anticipate drug success rates and potential side effects early on.
Expert-Dependent: Heavy reliance on domain experts for insights and direction.	**Reduced Expert Bias:** Can uncover novel insights without preconceived expert biases.
Empirical Screening: Uses high-throughput screening of large compound libraries against a biological target.	**In Silico Screening:** Employs virtual screening and predictive modeling, reducing reliance on physical compound libraries.
Cost Intensive: High costs due to lengthy trials and extensive manual labor.	**Cost-Efficiency Potential:** May lower costs by reducing the number of necessary physical experiments.
Regulatory Focus: Stringent regulatory processes tailored for traditional development pathways.	**Adaptive Regulatory Approach:** Emergence of new regulatory pathways that account for AI methodologies.

Defining the Multimodal Data

- Data from different sources that describes the same patient or research focus
    - Example:
      Patient-focused data from EHR, Pharmacy, Genomics, Proteomics, Patient-reported outcomes, imaging (radiology, nuclear medicine, etc.), pathology, etc.
- By using AI/ML techniques these data can be reviewed in the context of either supervised or unsupervised learning to find clusters and other findings suggestive of correlative findings



The Power of Hypothesis Generation

- Data Driven Insights
- Volume and Velocity
- Pattern Recognition
- Bias Reduction
- Integration of Multimodal data
- Iterative Learning

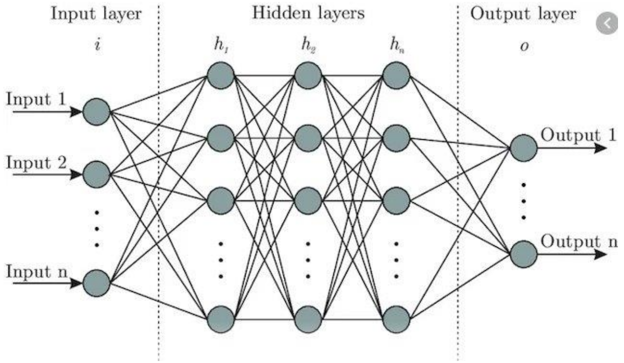
- Integration of multimodal data
- Unsupervised learning
- Scalability
- Resource optimization

AI can reduce the time needed analyze more data, faster, and more efficiently than in traditional hypothesis-driven drug discovery processes

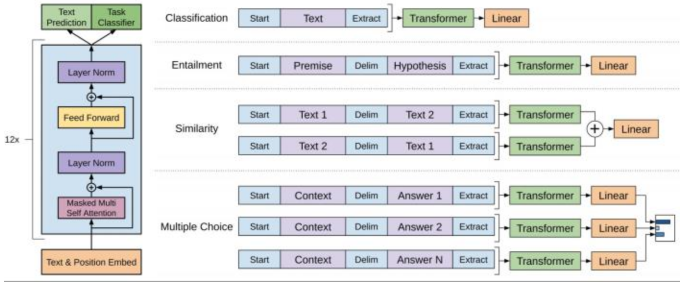
Key Technologies in AI for Drug Discovery

QUANTORI

- What is a Neural Network
- A Neural Network can be thousands of times more complicated than this drawing
- A GPT model (generative pre-trained model) can take thousands of tokens and simultaneously predict what’s next (in the case of LLMs)



1



2

Images from:

1. https://www.reddit.com/r/MachineLearning/comments/l1z8cr/d_best_way_to_draw_neural_network_diagrams/?rdt=33039
2. https://paperswithcode.com/method/gpt

Integrating Multimodal Data: Opportunities and Challenges

Performing AI-Learning on multimodal data can lead to:

- Comprehensive Understanding
- Improved Predictions
- Revealing Hidden Patterns
- Enhanced Biomarker Discovery
- Drug Repurposing:
- Target Identification
- Personalized Medicine
- Reducing False Positives/Negatives

The identification of a single biomarker can be the difference to a drug program that has regulatory success or failure. Using a single modality of data makes this far less likely.

Case Study:
Multimodal Data in Action

Identification of Drug Repurposing Targets

QUANTORI

Challenge

Our client was a biopharma company interested in repurposing a specific drug. They wanted to use multi-omics data from a large set of cell lines to prioritize targets for *in vitro* screening of drug activity.

Input data: gene expression levels, protein abundances, copy number variants, SNP genotyping.

Solution

- Dataset curation and pre-processing.
- Feature selection based on expert knowledge.
- Trained and evaluated ML models using different learning algorithms and feature configurations.
- Post-hoc interpretability analysis to determine key features used by the final model for prediction.

Benefits

- From >1000 candidate cell lines, the ML model was able to identify a set of ~80 high priority cell lines for *in vitro* screening.
- It is currently being deployed, with experimental screening results expected in Fall 2023.

Reduction in candidate cell lines from > 1000 to < 80 for screening.
Manual, single data set approaches could have taken YEARS

AI-Driven Drug Development

Challenge

Our early-stage pharma client was seeking to analyze datasets from multi-modal data types in the same domain that novel biomarkers and drug targets could be identified through AI algorithms. This approach could potentially reduce the need for a variety of clinical trials, which would create a strategic advantage for the client. But they lacked internal programming capabilities required to test their hypotheses.

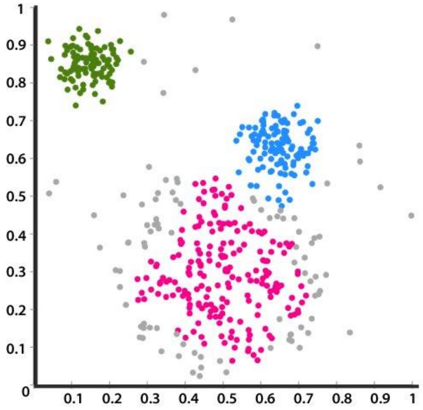
Solution

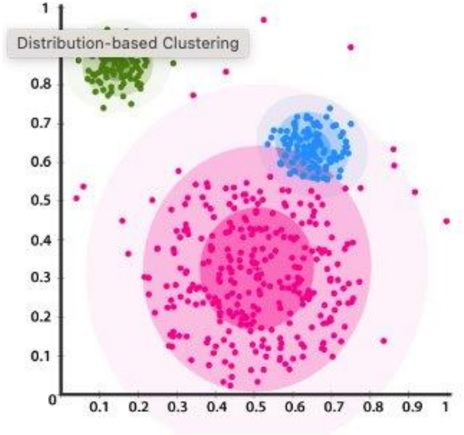
Using Quantori's workflows framework known as **QFlow**, our engineers created a series of tools and analytic pipelines to provide the company's data scientists with the ability to: 1) integrate all disparate data, 2) analyze that data for clustering and hypothesis testing and 3) provide rapid targeting capability to identify new biomarkers and test them on additional data sets.

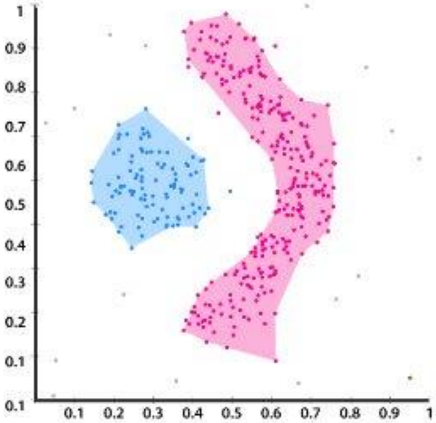
Benefits

- By deploying the Quantori accelerator, the client was able to leverage a valuable tool and additional capabilities in under 4 months, extending its financial runway.
- The client was able integrate, analyze and test its hypotheses real-time, resulting in new, key biomarker identification.
- The speed and increased throughput of these analyses led the firm to accelerate their development plans and consider filing for an IPO based on their ability to mobilize and analyze the data more efficiently.

AI-Driven Cluster Analysis







1

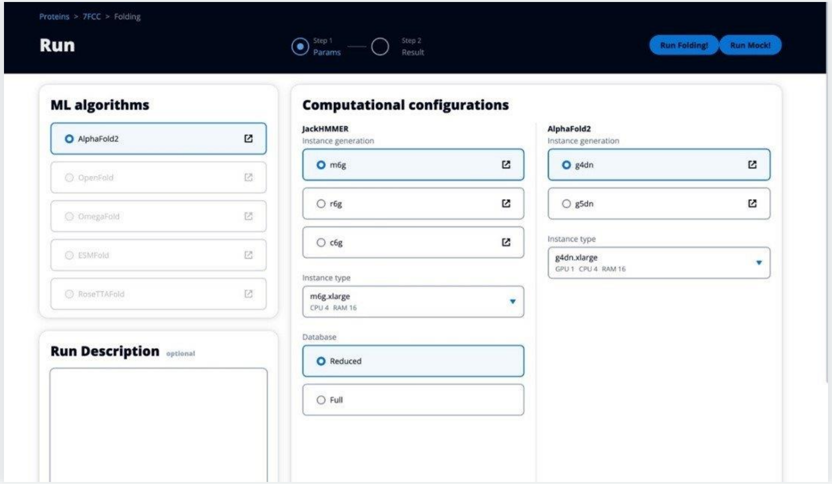
- Multimodal data can have clusters in "n" dimensions
- AI analysis can find clusters in this complex math space
- It can cut down analysis time

Images from:

1. https://byjus.com/maths/cluster-analysis/

AI-Driven Drug Design

- Protein structure prediction was previously computationally time-consuming, expensive, and had accuracy issues.
- AlphaFold is an application that does these calculations and modeling requiring virtual machines and complex setup.
- Quantori developed a novel AI-driven interface to interact with AlphaFold that streamlines and speeds up business processes.
- Leveraging AWS Batch from Jupyter Notebook interfaces, our tools split the tasks, assigning those optimized for CPU vs. GPU appropriately – speeding up processing times.
- RAG Status – a series of traffic lights are available in the tool to cue the users to costs vs time trade offs.



ML algorithms
- ● AlphaFold2
- ○ OpenFold
- ○ OmegaFold
- ○ ESMFold
- ○ RoseTTAFold

Computational configurations

JackHMMER

Instance generation
- ● m6g
- ○ r6g
- ○ c6g

Instance type: m6g.xlarge CPU 4 RAM 16

Database
- ● Reduced
- ○ Full

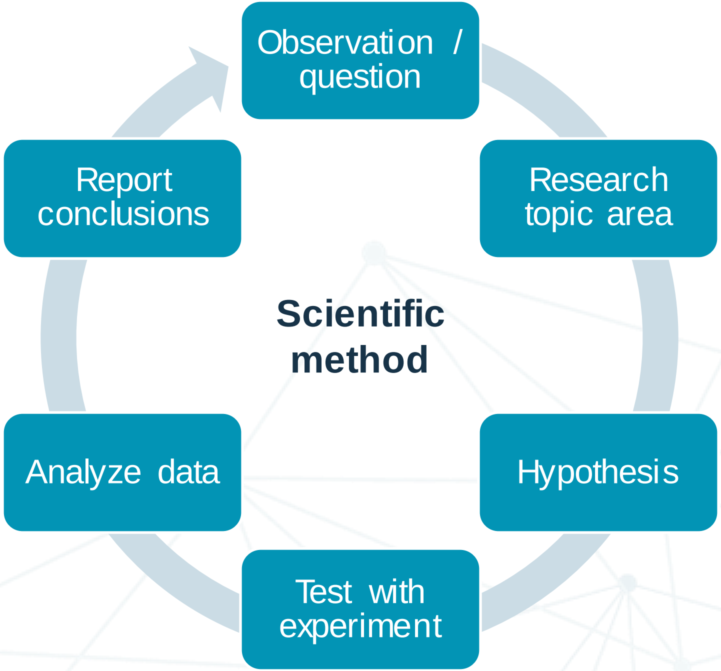
AlphaFold2

Instance generation
- ● g4dn
- ○ g5dn

Instance type: g4dn.xlarge GPU 1 CPU 4 RAM 16

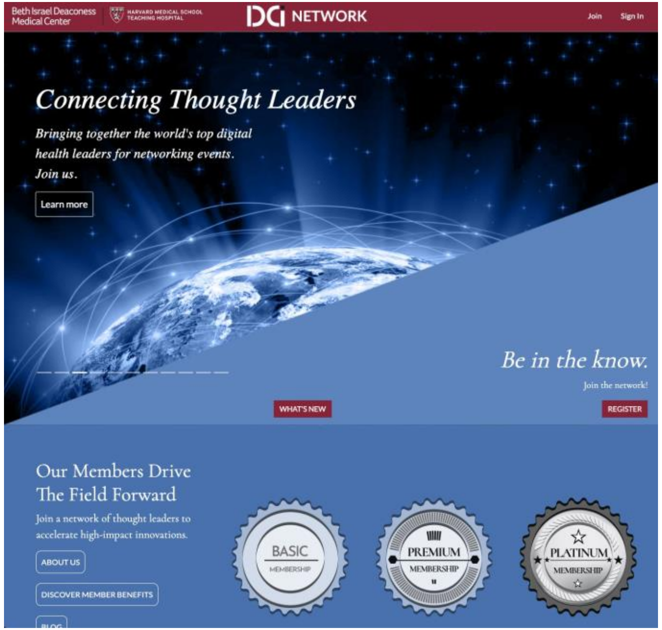
Hypothesis Generation: From Theory to Practice

- AI and machine learning puts the scientific method on steroids
- It speeds it up
- It makes it more accurate
- It can find clusters to be tested far faster and efficiently than in past generations



Division of Clinical Informatics, Harvard Medical School

- A Harvard-based think tank
- Draws from a multistakeholder group of stakeholders
  - Academics
  - Business
    - Life Science
    - Technology Companies
  - Government/Regulatory Agencies
  - Patients
- 3 Foci
  - Patient Engagement in Healthcare via Medication Labels
  - Collaborative Precision Oncology
  - AI in Healthcare



Connecting Thought Leaders

Bringing together the world's top digital health leaders for networking events. Join us.

Be in the know.

Join the network!

Our Members Drive The Field Forward

Join a network of thought leaders to accelerate high-impact innovations.

www.DCINetwork.org – www.DCINetwork.org/events

Introduction to AMIA (American Medical Informatics Association)

QUANTORI

- 5500 Medical and Healthcare Professionals
- Focused on Computer applications in Healthcare
- Over 20 WGs
- Latest focus in collaboration with the DCI Network was
  - AI in Healthcare
  - Collaborative Precision Medicine
  - Patient Engagement around Medical Information
- Three Major Conferences:
  - Annual Symposium (November)
  - The Bioinformatics Summits (March)
  - The Clinical Informatics Conference (May)
- Multiple Journals
- Focus with DCI Networks on AI in Healthcare
- Drug Development is one major focus of AMIA

AMIA

INFORMATICS PROFESSIONALS. LEADING THE WAY.

Navigating the Challenges in AI for Drug Development

Multimodal data analysis for drug discovery represents a cutting-edge approach that integrates diverse types of data, including genomic, proteomic, clinical, and chemical data, to uncover novel insights that can accelerate the development of new drugs. However, this approach also comes with several significant challenges:

- **Data Heterogeneity**
- **Data Volume and Complexity**
- **Data Quality and Incompleteness**
- **Integration and Analysis Techniques**
- **Interpretability and Explainability**
- **Regulatory and Ethical Considerations**
- **Collaboration and Data Sharing**

Despite the complex challenges, this remains the cutting edge for discovery and drug development

Promising Trends in Multimodal Data Analysis

QUANTORI

The Issues	Comments
Advanced AI and Machine Learning Models	We are literally at the very beginning of this AI revolution. New techniques are likely to leapfrog
Improved Data Integration Techniques	New approaches to using LLMs or other AI/ML techniques will help bring new data sets on line faster and more efficiently
Enhanced Interpretability and Explainability	Trust in the system continues to be an issue. The sooner we get to testing for trust, the faster we will advance the space
Emphasis on Real-world Data and Patient-centric Approaches	RWD continues to be at the forefront of this space – when merged with other data sets, this holds the key for new drug uses and discoveries – when data quality improves
Cross-disciplinary Collaborations	Its not enough to have just data scientists at the table – you need SMEs (Clinical Informaticians, Chemists, etc) to advance
Ethical AI and Responsible Data Use	Biases in the data training sets is inevitable. We must provide means to ensure that either we call out biases, or we work to ensure transparency
Expansion of Digital Biomarkers	When merging datasets for multimodal analysis, a new kind of biomarker is emerging, “in silico” biomarkers. Found by using multimodal analytics

Acknowledgments

DCi NETWORK

www.DCINetwork.org

AMIA

INFORMATICS PROFESSIONALS. LEADING THE WAY.

www.AMIA.org

Contact Information

Steven E. Labkoff, MD, FACP, FACMI, FAMIA

Global Head, Clinical and Healthcare Informatics

steven.labkoff@quantori.com

+1 (917) 599-7742