AI in Target and Biomarker Discovery

Understanding disease biology using AI to enhance R&D efficiency

July 2022

Advisors

MEHTA PARTNERS

Executive summary

Artificial Intelligence can augment the development of disease models to offer a more informed hypothesis support for drug discovery and development

- AI allows the processing of volume-velocity-variety of complex scientific data in effective multi-dimensional models
- Advent of new technologies (large scale omics profiling, knowledge graphs, etc.) presents a significant opportunity not only to de-risk the programs but also generate new ideas.

AI can extract hidden signals to build effective hypothesis for target/biomarker identification, repurposing, patient stratification and more

- AI companies use a variety of data types OMICs, unstructured text data (publications, grants, patents), databases, EHR data, pathology data or a combination of multiple data types to drive outcomes with services or self-use software's

BioPharma AI space is gaining momentum with several tech companies offering cutting edge technologies and increasing partnerships with a keen focus on target/biomarker identification

- Since 2011, a total of ~350 AI companies have been incorporated addressing various facets of drug discovery and development

A well-rounded assessment with several considerations are critical to select the right tools/partners from the pool of options

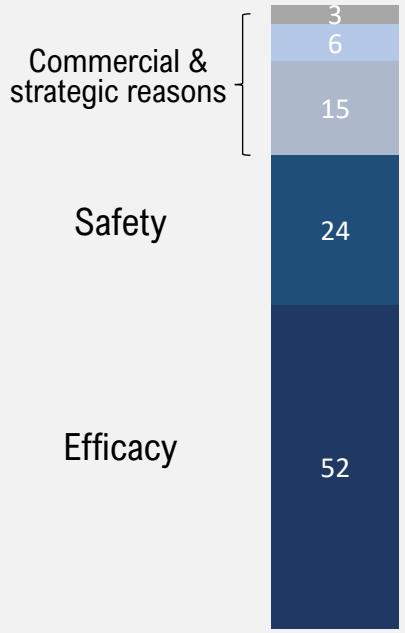
- Defining the specific business problem to solve, checking the cleanliness of data for application of ML, selecting the right machine learning model, understanding nuances of operations, security etc. are important to create a successful AI strategy

With over 3 decades of diverse and global biopharma experience and deep understanding of the AI space, MP Team can effectively catalyze your AI initiatives

Most drugs fail because biology is complex

Spread of data across disparate sources and ever-expanding literature leads to inefficient hypothesis building

Understanding the role of genes/targets in disease biology is a crucial step in drug discovery and can define the outcome in terms of the success or failure of trials. However, the outcome of choice becomes fully obvious only years later during clinical development.



Underlying reasons

- Stopped for the company's change in strategy.
- Unknown effects due to loss of protein-protein interactions.
- Drug also binds to OFF-targets leading to toxicity.
- Pathway has other 'Backup' proteins and may not alter the disease enough.
- Intervention is difficult as the drug is unavailable at the target site.

Reasons for failure in phase II and III clinical trials (2013-2015)

Advent of new technologies present a significant opportunity to de-risk the programs:

Genomics/Large scale profiling of individuals to understand the association of genes with a wide variety of diseases.

Transcriptomic and Proteomics data analysis to understand the change in expression of target proteins during different stages of the disease.

Years of clinical and literature data are being organized into portals and searchable databases.

However, traditional approaches are unable to take a multi-factorial approach to connect and synthesize complete insights from signals spread across massive yet diverse data types.

Source: Harrison (2016). *Reasons for clinical failures analysis.* Nat Rev Drugs Discov.

Artificial Intelligence can help develop better disease models

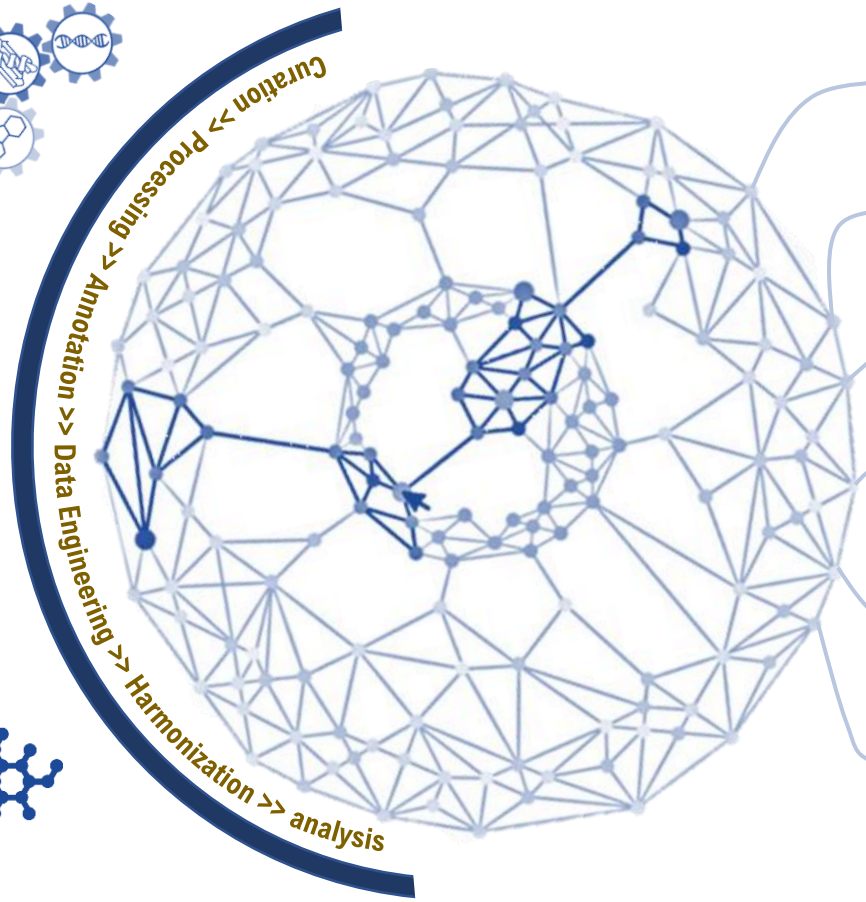
AI allows the processing of volume-velocity-variety of complex scientific data in effective multi-dimensional models

OMICS data:
Genomics, proteomics, transcriptomics, metabolomics

Unstructured text data:
Publications, books, patents, etc.

Clinical trial data:
Trial outcomes, EHR, pathology images

Molecule libraries:
Structure-Activity Relationship, Tox datasets



Map novel disease pathways combining multi-omics and literature data to identify novel targets.

Predict future toxicity/efficacy using predictive models.

Identify distinct subtypes of the disease with different clinical manifestations and possibly different molecular root causes.

Identify new links between unknown targets and existing drugs for repurposing opportunities.

Precision biomarker discovery to select the right patients for targeted clinical trials.

Link genomic data and identify biomarkers to monitor drug response or progression of the disease.

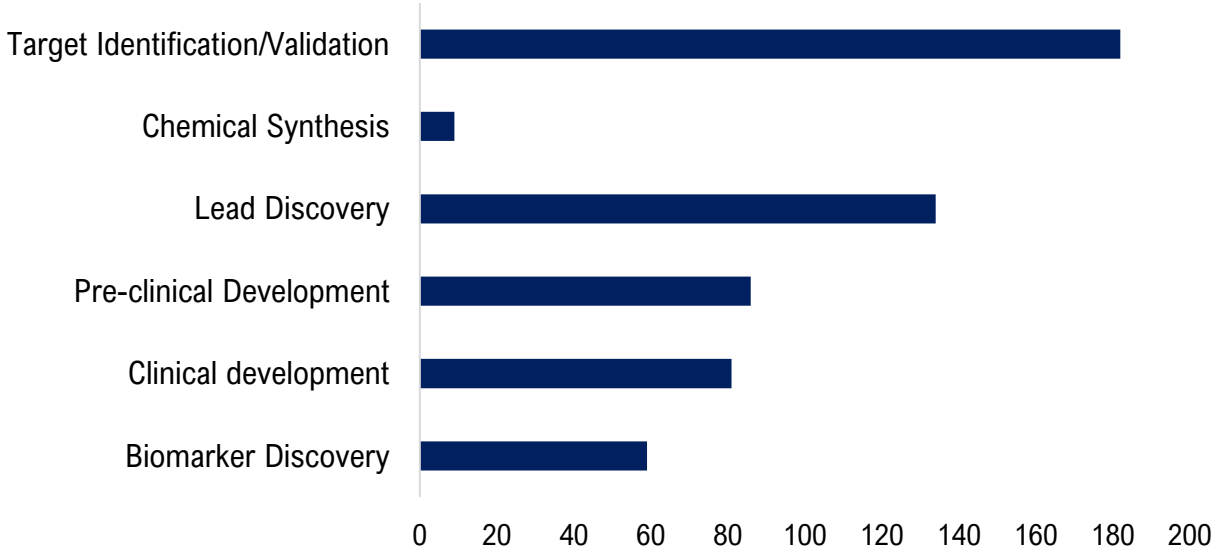
BioPharma AI space is gaining momentum

Target identification/validation and biomarker discovery remain one of the top research use cases for AI

Since 2011, the AI landscape has seen a >600% increase in the number of companies with the total of ~350 companies in 2021.

Target and biomarker discovery are among the key areas of research cases being addressed by these technologies followed by small-molecule drug design.

Distribution of AI companies by stages in the lifecycle



Source: BiopharmaTrend, News, MP analysis

2019 — Deep genomics uses large scale OMICs analysis to identify a new target and oligonucleotide therapy for Wilson's disease in just 18 months.

2020 — Benevolent AI in collaboration with AstraZeneca identifies and validates novel drug targets for CDK and IPF using a knowledge graph approach.

2020/21 — CytoReason develops holistic disease model (OMICs, text/KG, clinical data) and partners with Pfizer, Ferring, Sanofi, Roche and Merck KGaA for target id and indication selection.

2021 — Owkin and Amgen predictive AI model, based on EHR data, outperforms traditional methods to predict heart attack and stroke.

2022 — BMS, Roche and PathAI collaborate for digital image analysis (CD8 biomarker analysis based on pathology images).

A few key milestones/partnerships in the last 3-4 years

AI can catalyze most traditional workflows and applications

Multi-modal approaches provide robust hypothesis generation through multiple evidence sources

Data source	Increasing complexity		
 OMICs	▪ Seq analysis (DNA seq, RNA seq, Proteomics) to find even weak signals.	▪ Identify compact and explainable biomarkers. ▪ Automated data curation and meta-annotation.	▪ Pathway analysis towards understating MOA.
 OMICs + Text	▪ All above analyses augmented with literature support.	▪ Comprehensive pathway analysis with augmented support from key biology databases.	▪ Building complete in-silico disease models for predicting drug toxicity/efficacy by simulations.
 Text	▪ Extract key insights from thousands of publications/clinical trials in a structured data table for analysis.	▪ Build hypothesis around MOA/target id/repurposing by exploring data from diverse sources for a structured review.	▪ Identify hidden relationships between drug-disease-targets for which enough implicit data may not exist yet.
 Pathology + EHR data ± OMICS	▪ AI driven analysis of pathology images is currently being deployed to identify biomarkers for patient stratification. This is often used in conjunction with ML/NLP driven analysis of EHR reports and sometimes with OMICs data to drive biomarker id/stratification.		

Diverse AI players solving similar problems with different offerings

Complexity of the application typically guides the business model

Business models

Data source	SaaS	HYBRID (SaaS first)	HYBRID (Services first)	SERVICES
 OMICs	Data4Cure, InSyBio	DeepLife., lucidata	PHENOMIC.AI	OCEAN GENOMICS, Era7 bioinformatics
 OMICs + Text	pandaOmics, EURETOS	Biorelate, ROSALIND	Biovista	biotx.AI, Benevolent AI
 Text	Plex, PERCAY AI	Linguamatics, causaly	Wisecube	Data2Discovery, Melax Tech
  Pathology + EHR data ± OMICS	nucleai, aiforia	Genialis, PAIGE	PathAI	Ardigen, Ariana pharma

Case study

elucidata

Founded

2013

Location

United States

Size

131 employees

Funding

$7.6 M (pre-series A)

Key Clients

Key Offerings: Curated OMICs data atlases; Automated OMICs data curation pipelines; Configurable ML-Ops platform.

Elucidata's platform, Polly, is an ML Ops platform with several predefined pipelines for OMICs data and network analysis. Additionally, they offer meta-annotated and clean omics data atlases that can be accessed through APIs or used for analysis on their platform.

Problem

- To identify differentiation targets in Acute Myeloid Lymphoma or AML and other oncology indications such as neuroblastoma and melanoma.
- Development of accurate, explainable ML models for Target ID and Patient stratification.

Solution

- Made OmixAtlas available across enterprise enabling Auron's internal research team to collaborate.
- Built a customized pipeline to process heterogenous data (public and in-house).
- Built and deployed a proprietary ML model for patient classification.

Impact

- The application of Polly platform helped Auron identify 2+ novel targets in AML, of which 1 was validated.
- The target was identified within 2-3 months, significantly shorter than the average 1-2 years time period.

OMICS data: Genomics, transcriptomics, metabolomics

Unstructured text data, Publications, books, patents, etc.

Clinical trial data: Trial outcomes, EHR, pathology images

Molecule libraries: Structure-Activity Relationship, Tox datasets

Source: Company website; Primary research

Case study

Genialis

Founded
2011

Location
United States

Size
29 employees

Funding
$2.5 M (Seed round)

Key Clients

Roche
OncXerna THERAPEUTICS
Boehringer Ingelheim
Sutter Health

Key Offerings: Biomarker discovery platform; Software for analyzing sequencing data; ML models for patient stratification.

Genialis' platform, ResponderID, provides a data management framework to organize and analyze OMICs datasets for biomarker id and patient stratification. Additionally, they offer their proprietary software, Expressions with pre-configured pipelines to interpret and visualize sequencing data.

Problem

- To turn Xerna's population-based TME panel into a disease-agnostic predictive biomarker system.
- Development of accurate, clinically robust ML models for regulatory and commercial purposes.

Solution

- Artificial Neural Network applied on RNA-seq data from patient tumours to classify patients into one of the stromal phenotypes of cancer.
- Biomarker status is decided based on the mechanism of action or MoA of each drug.
- Non-population-based model ready for clinical development.

Impact

- The Xerna TME Panel now includes normalized gene expression data.
- The Xerna TME Panel is in development as a clinical trial assay and has been licensed by Qiagen for development as a companion diagnostic for navicixizumab as a Research Use Only assay.

OMICS data: Genomics, transcriptomics, metabolomics

Unstructured text data, Publications, books, patents, etc.

Clinical trial data: Trial outcomes, EHR, pathology images

Molecule libraries: Structure Activity Relationship, Tox datasets

Source: Company website; Primary research

Case study

Benevolent AI

Founded

2013

Location

United Kingdom

Size

342 employees

Funding

$292 M (IPO date: Apr25, 2022)

Key Clients

EpiGen Global Research Consortium

NOVARTIS

AstraZeneca

Sano Genetics

Key Offerings: A knowledge graph-driven platform built on scientific literature and relevant biomedical data.

BenevolentAI offers a knowledge graph (KG) driven platform to identify novel targets, enable precision medicine, and support the design of molecules.

Problem

- To address the rapid spread of 2019-nCoV, a drug needs to be identified/repurposed before the availability of vaccines.
- Development of accurate ML models to mine through the scientific literature.

Solution

- BenevolentAI's KG-based platform identified 47 medically approved AAK1 inhibitors of the available 378 with 6 candidates showing high affinity.
- The platform further refined its search based on side-effects of the drugs thereby suggesting baricitinib to be trialed on an appropriate patient population.

Impact

- Patients with bilateral COVID-19 pneumonia were treated with barcitinib.
- The treatment showed a reduction in SARS-CoV2 viral load and inflammatory markers.
- The data supported further evaluation in randomized trials.

OMICS data: Genomics, transcriptomics, metabolomics

Unstructured text data, Publications, books, patents, etc.

Clinical trial data: Trial outcomes, EHR, pathology images

Molecule libraries: Structure-Activity Relationship, Tox datasets

Source: Company website; Primary research

Case study

PathAI

Founded

2016

Location

United States

Size

345 employees

Funding

$255.2 M (Series C)

Key Clients

GlaxoSmithKline

AKOYA BIOSCIENCES

Cleveland Clinic

Roche

Key Offerings: AI-powered technology for pathology, diagnostics, and predictive molecular signatures.

PathAI offers platforms and services based on pathology analyses thereby improving the accuracy of diagnosis and treatment efficacy.

Problem

- To identify novel tumor micro-environment or TME features in a Phase 3 study (IMpower150) for NSCLC.
- Development of ML models to understand pathological signatures of VEGF inhibition in the context of immuno-oncology.

Solution

- A deep CNN developed using H&E images digitized on the platform.
- Applied 100k annotations from expert pathologists to segment tissue by area and label individual cell types.

Impact

- Identified the relation between high LFR (ratio of lymphocyte to fibroblasts) and improved PFS (progression-free survival).
- The effect of TME and vasculature on PD-L1 and VEGF-targeting therapies.

OMICS data: Genomics, transcriptomics, metabolomics

Unstructured text data: Publications, books, patents, etc.

Clinical trial data: Trial outcomes, EHR, pathology images

Molecule libraries: Structure Activity Relationship, Tox datasets

Source: Company website; Primary research

Key considerations prior to AI adoption

A well-designed pilot study, while setting realistic expectations is critical while assessing the platform and gain long term confidence

With several companies and tools addressing similar problems, it is crucial to invest time and resources to identify and assess the right tool for one's needs. Understanding internal capabilities/skill sets and priorities are equally important while assessing the platform to establish the right fit. A poorly designed collaboration is likely to yield disappointment due to unrealistic expectations.

Problem statement

Is the adoption to improve the overall efficiency or accuracy? Or are you trying advance a specific scientific problem? Is the aim to empower/build an internal team for the future or identify an outsourcing partner?

Data

Do you have enough data to apply ML/DL? Is it public or private data? Is the data clean? Does the partner have the right skill sets to clean the data? Does the partner have proprietary data sets?

Models, benchmarking and modularity

The selection of the right type of ML (supervised, unsupervised, active, transfer, etc.) is critical for the functioning of the platform. Does the partner have the right models? Are the benchmarking studies done with a suitable dataset? Does the platform allow integration of your internal models?

Operations and Security

Would the data stay on your cloud/premise or be transferred to the partner's ecosystem? Is the system secure? When would the client delete data from the cloud? Is it enough time to revisit the analysis?

Partnership structure

Is it simple enough to adopt as a SaaS after initial training? Does the partner provide implementation services? How to design a successful POC with enough stage gates?

Return on Investment

Is it an out-of-the-box solution or will need testing before providing the desirable outcome? How to model the ROI? What internal skill sets can improve ROI?

MP Group can catalyze your AI initiative

MP group is deeply involved in the AI for pharma space. We regularly interact with 50+ AI biotechs offering diverse AI-driven applications in the drug discovery lifecycle.

With over 3 decades of diverse experience and integrated perspective in domestic and global BioPharma, MP Team will be happy to be an extension of your management team and help with one or more of the below initiatives:

- Asses the internal capabilities and identify the key business segments for potential disruption/augmentation by AI platforms
- Identify business segments for short-term and long-term benefits from AI interventions
- Identify partnering or investment opportunities unique/relevant to the company's vision
- Technical due diligence to investigate the AI platforms best suited for the specific needs
- Build and implement short term and long-term AI strategy for drug discovery

We invite you to write to us -

Viren Mehta
mehta@mpglobal.com

Neel Fofaria
neel@mpadvisor.com

Amandeep Singh
amandeep@mpadvisor.com

MEHTA PARTNERS

Advisors