Drugcode Biotech
The next-generation of drug prediction platform

Hongyu Zhao, Wajahat Mehal, Jerome Yu, Xinshou Ouyang
DrugCode Team

Hongyu Zhao, PhD
Ira V. Hiscock Professor of Biostatistics, Professor of Genetics, Professor of Statistics and Data Science of Yale University

Wajahat Mehal PhD, MD, BA
Professor of Medicine (Digestive Diseases), Director of Yale Weight Loss Program, Director of Yale Fatty Liver Disease Program

Zhaolong (Jerome) Yu, MS
PhD Candidate in Computational Biology and Bioinformatics of Yale University

Xinshou Ouyang, PhD
Assistant Professor of Yale University
DrugCode Biotech

Our Product:
• A validated multiomics platform for drug discovery and prioritization

Competitive Edges:
• Novel statistical methods for drug prioritization
• Comprehensive integration of disease multiomics datasets and cell-type specific drug-induced perturbation databases
• Computational prediction and experiment validation

Business Models:
• In-house drug discovery and pipeline building
• Drug prioritization service for Pharma and Biotech

Steps Required to Secure Funding:
• Commercial software development
• Large-scale perturbation database building
• In-vivo studies for testing NASH and prostate cancer drug candidates
Unmet Need: *de novo* R&D is risky and omics resources are not fully used

**Huge markets**

<table>
<thead>
<tr>
<th>Market</th>
<th>Size (B) (CAGR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolic</td>
<td>$61.10 (7.56%)</td>
</tr>
<tr>
<td>Oncology</td>
<td>$128.45 (6.83%)</td>
</tr>
<tr>
<td>Organ Fibrosis</td>
<td>$3.21 (6.08%)</td>
</tr>
</tbody>
</table>

**Limitations of traditional drug discovery**

- **Time-consuming**
  - The process takes 10-15 years
- **Huge capital investment**
  - $2.56 B for a new drug
- **Low success rate**
  - 1 out of 10,000 – 15,000 compounds

**Traditional R&D process**

<table>
<thead>
<tr>
<th>Phase</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discover and Development</td>
<td>6.5 Years</td>
</tr>
<tr>
<td>Preclinical Research</td>
<td></td>
</tr>
<tr>
<td>Clinical Research</td>
<td>7 Years</td>
</tr>
<tr>
<td>FDA Review</td>
<td>1.5 Years</td>
</tr>
</tbody>
</table>

**Large-scale omics resources are not fully exploited**

- **Genome-wide association studies**
  - Associated loci are in non-coding region
- **Large-scale drug-induced perturbation database**
  - Measures the effects of compound treatment
  - LINCS database (21,231 small molecules, 88 cell lines)

Source: Saint John’s Cancer Institute
Overview of the DrugCode platform
A Multiomics platform for drug discovery and prioritization

DrugCode drug development platform
- built on large-scale disease genetic data and drug-induced perturbation databases

- DrugRank™
  - Drug Perturbation Database
  - Multiomics Scoring Method
  - In-house Pipeline
  - Pharma
  - Research Partners
  - Drug Candidates

- DrugSpec™
  - Compounds with Highest Scores
  - In Vitro Efficacy Report

DrugCode platform solution highlight

Fast and Economical
- Complementing early studies

De-risking
- Identifying high success targets from a large pool

High Success Rate
- Better efficacy and safety profiles

Non-disruptive
- Works with established drug discovery pipelines

Source: self-analysis.
Methodology of the DrugCode platform
A Multiomics platform for drug discovery and prioritization

Comprehensive Transcriptome Analysis
- Differentially expressed genes (DEG)
- Case/Control DEGs
- Drug intervention DEGs
- Reversible signatures

Disease Gene Identification via GWAS and TWAS
- Genome-wide association studies (GWAS) of disease-associated traits
- Transcriptome-wide association studies (TWAS) based on GWAS summary statistics *

Prior Disease Pathway Knowledge
- Prior knowledge about target pathways
- Combined signatures

KS-statistics-based Drug Prioritization

We will show results for two very different conditions: Non-alcoholic Steatohepatitis and Prostate Cancer

## Competitive Landscape

<table>
<thead>
<tr>
<th>Competitors</th>
<th>Omics Data Integration</th>
<th>Wet Lab Capabilities</th>
<th>Scalability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DrugCode Biotech:</strong> Multiomics computational drug discovery and prioritization</td>
<td>• Transcriptome Pathway analysis GWAS TWAS</td>
<td>• Yes</td>
<td>• Yes</td>
</tr>
<tr>
<td><strong>Lantern Pharma:</strong> (NASDAQ) R&amp;D for in-licensing drugs based on drug efficacy markers and machine learning</td>
<td>• Transcriptome EHR</td>
<td>• No</td>
<td>• No</td>
</tr>
<tr>
<td><strong>Standigm:</strong> (pre-IPO $44.5 M) full-stack AI-driven industrializing drug discovery</td>
<td>• Knowledge Graph</td>
<td>• Yes</td>
<td>• Yes</td>
</tr>
<tr>
<td><strong>Dr. Jinghua Gu’s Lab:</strong> Signature-free drug repositioning method Dr. Insight</td>
<td>• Transcriptome</td>
<td>• No</td>
<td>• Yes</td>
</tr>
<tr>
<td><strong>Dr. Yadong Huang’s Lab:</strong> Computational drug repurposing with genotype-dependent RNA signatures</td>
<td>• Transcriptome GWAS</td>
<td>• Yes</td>
<td>• No</td>
</tr>
<tr>
<td><strong>Lantern Pharma:</strong> (NASDAQ) R&amp;D for in-licensing drugs based on drug efficacy markers and machine learning</td>
<td>• Transcriptome EHR</td>
<td>• No</td>
<td>• No</td>
</tr>
<tr>
<td><strong>Standigm:</strong> (pre-IPO $44.5 M) full-stack AI-driven industrializing drug discovery</td>
<td>• Knowledge Graph</td>
<td>• Yes</td>
<td>• Yes</td>
</tr>
<tr>
<td><strong>Dr. Jinghua Gu’s Lab:</strong> Signature-free drug repositioning method Dr. Insight</td>
<td>• Transcriptome</td>
<td>• No</td>
<td>• Yes</td>
</tr>
<tr>
<td><strong>Dr. Yadong Huang’s Lab:</strong> Computational drug repurposing with genotype-dependent RNA signatures</td>
<td>• Transcriptome GWAS</td>
<td>• Yes</td>
<td>• No</td>
</tr>
</tbody>
</table>

Source: team analysis; lit-search
Results I: Drugs in non-alcoholic steatohepatitis (NASH) pipeline are significantly enriched in the Top 5% and 10% drug recommendation list

Experiment Setup

1. Ranking the drug list based on previously identified comprehensive disease signatures
2. Testing the percentages of known drug candidates in the predicted drug list.

Started with a database of 5774 drugs containing 30 drugs known to be in the NASH pipeline. Our analysis predicted the top 5% (289) most likely NASH drugs, and 13/30 known NASH drugs were in this top 5%. Our analysis took a database containing 30/5774 (0.5%) NASH drugs and enriched to 13/289 (4.5%). Most drugs are not currently in a NASH pipeline.
Result II: in vitro efficacy validation for top seven NASH drug candidates in mouse primary hepatocytes made steatotic with palmitate

Success rate is high
- Among the top 7 available compounds, 4 of them showed efficacy of reducing the lipid level in the mouse primary hepatocytes.
- None of these 4 compounds are currently in public NASH drug pipelines.
Business model: in-house pipelines and strategic partnership

1. Build In-house Pipelines to Identify NCE from Available Databases

- **Disease areas**
  - Metabolic disease (NASH)
  - Oncology (Prostate cancer)
  - Fibrosis

- **Translational Research**
  - Collaboration within research institutes and hospitals to identify potential drug candidates

- **Compounds**
  - Screen from currently available perturbation database
  - Screen from in-house perturbation database

2. Strategic Partnership with Pharma and Biotech

- **Drug prioritization**
  - De-risking the R&D efforts by screening for the most promising candidates

Source: self-analysis.
Summary of and Unique Aspects of Product

Highlight

A. Tested Multiomics Drug Prediction Method
A proprietary way to combine drug modified gene expression data, GWAS data and TWAS data to make predictions about the ability of molecules to have efficacy in disease states, in-combination with testing of predictions using in-vitro and in-vivo multiple models.

B. *In Silico* Target Pathway Evaluation
Using CRISPR or shRNA screening libraries to assess the importance of the target pathway in terms of potential drug efficacy

Use Cases

A: Identify New therapeutic candidates.

B: Prioritization at the level of pre-screened molecules.

C: Prioritization at the level of the target.

Source: self-analysis.
Use of Blavatnik Funds: 300K over two years

Platform Development

1. DrugCode commercial software development and implementation (50K)

2. Build in-house perturbation database from small molecule library (100K)
   • Microarray assays to measure the expression profiles of disease relevant cell lines after treatment with small molecules.

In-house Pipeline Development

1. Allow testing of current candidates in multiple models of NASH and Prostate Cancer (100K)
   • The top 5 hits from the NASH screen will be used in two in vivo models. These will be the Western Diet and the Choline Deficient Amino Acid supplemented models.
   • The Top 5 hits from the prostate cancer screen will be used in two in-vivo models of prostate cancer.

2. Apply the DrugCode platform to predict and in-vitro test molecules with anti-fibrotic efficacy (50K)
   • Therapies for organ fibrosis are a huge unmet need, and a large amount of expression and GWAS data is available to use with the DrugCode platform. We will initially start with lung and liver fibrosis predictors and then test the top hits in vitro.
Development Plan and Timeline

<table>
<thead>
<tr>
<th>Activity</th>
<th>2021</th>
<th>2022</th>
<th>2023</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q4</td>
<td>Q1</td>
<td>Q2</td>
</tr>
<tr>
<td>Team building</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fund raising</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software development</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perturbation database building</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NASH in vivo study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prostate cancer in vivo study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fibrosis drug development</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: self-analysis.

10/15/21

Blavatnik Funding Start
Thank you!

Drugcode Biotech
The next-generation of drug prediction platform