Machine Learning for Prediction of Cancer Drug Response

Rick Stevens
Argonne National Laboratory
The University of Chicago

Crescat scientia; vita excolatur
CANcer Distributed Learning Environment (CANDLE)

DOE-NCI partnership to advance exascale development through cancer research

Rick Stevens and Tom Brettin
Argonne National Laboratory
University of Chicago

“... But the true method of experience on the contrary first lights the candle, and then by means of the candle shows the way; commencing as it does with experience duly ordered and digested, not bungling or erratic, and from it educing axioms, and from established axioms again new experiments;”

—Francis Bacon in Novum Organum

August 28, 2018

Presented to:
ECP AD KPP Review
The NCI-DOE Partnership is Extending the Frontiers of Precision Oncology (Three Pilots)

• **Cancer Biology**
  – Molecular Scale Modeling of RAS Pathways
  – Unsupervised Learning and Mechanistic models
  – Mechanism Understanding and Drug Targets

• **Pre-clinical Models**
  – Cellular Scale PDX and Cell Lines
  – ML, Experimental Design, Hybrid Models
  – Prediction of Drug Response

• **Cancer Surveillance**
  – Population Scale Analysis
  – Natural Language and Machine Learning
  – Agent Based Modeling of Cancer Patient Trajectories
CANCER

1 in 4 of us will die from CANCER

Medical Sure presents an infographic look at

1 in 4 people in the USA will die from cancer

Although this rate has been dropping consistently since the peak in 1991, when 215 people died from one of cancer or another per 100,000 US residents.

In the United States, the cancer death rate of:

59% in 2009

Cancer death rates (USA):

- Heart and soft tissue: Death Rate: 42%
- Brain and nervous system: Death Rate: 70%
- Digestive system: Death Rate: 62%
- Lung and respiratory system: Death Rate: 80%
- Eye and orbit: Death Rate: 10%
- Bones and joints: Death Rate: 48%
- Prostate and male genital system: Death Rate: 14%
- Female genital system: Death Rate: 32%

Approximately 1500 people die every day from a form of cancer.

www.medicalsure.co.uk
What Is CANCER?
What is Cancer?

• Large number of complex diseases
• Each behave differently depending on cell type from which originate
  • Age on onset, invasiveness, response to treatment
• Common General Properties
  • Abnormal cell growth/division (cell proliferation)
  • Malignant tumors
  • Spread to other regions of body (metastasis)
Normal cell → Cell divides to create organized structure → Mutation in cell creates abnormal growth

- Cell is destroyed by self-inflicted cell death or immune system
- Cell bypasses death signals and continues to grow

Restores normal structure → Cancer forms
Acquired Capabilities of a Cancer Cell

- Self-sufficiency in growth signals
- Evading apoptosis
- Insensitivity to anti-growth signals
- Sustained angiogenesis
- Tissue invasion & metastasis
- Limitless replicative potential

Cell 100, 57–70
Mutations that Change Cell Behavior

Match Normal Pairs (GDC) showing translation
In Gene Expression Feature Space

Hanahan & Weinberg
Cell 100:57 2000
<table>
<thead>
<tr>
<th>Drug and Indication</th>
<th>Cost per Year of Treatment, $*</th>
<th>Parent Drug</th>
<th>Mechanism of Action</th>
<th>Clinical Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sorafenib for papillary thyroid cancer</td>
<td>140 984</td>
<td>NA</td>
<td>First approved VEGF and RAS tyrosine kinase inhibitor</td>
<td>Median PFS, 10.8 vs 5.8 mo</td>
</tr>
<tr>
<td>Capmatinib for in-situ skin cancer</td>
<td>17 672</td>
<td>NA</td>
<td>Anti-HER2 tyrosine kinase inhibitor</td>
<td>Median OS, 7.6 vs 3.9 mo</td>
</tr>
<tr>
<td>Lenvatinib for advanced renal cell cancer</td>
<td>18 124</td>
<td>A</td>
<td>Block the renal cell growth factor</td>
<td>RR, 18% vs 11%</td>
</tr>
<tr>
<td>Olmutuzumab for chronic lymphocytic leukemia</td>
<td>74 304</td>
<td>Rituximab</td>
<td>Anti-CD20 monoclonal antibody</td>
<td>Median PFS, 21.0 vs 11.0 mo</td>
</tr>
<tr>
<td>Pembrolizumab for breast cancer</td>
<td>78 252</td>
<td>Pembrolizum</td>
<td>Anti-PD-1 monoclonal antibody</td>
<td>Pathologic CR 39.3% vs 21.5%</td>
</tr>
<tr>
<td>Nab-paclitaxel for breast cancer</td>
<td>60 000</td>
<td>Paclitaxel</td>
<td>Block the microtubule inhibitor</td>
<td>Median OS, 9.5 vs 6.1 mo</td>
</tr>
<tr>
<td>Afatinib for non-small-cell lung cancer</td>
<td>79 920</td>
<td>Erlotinib</td>
<td>EGFR tyrosine kinase inhibitor</td>
<td>Median PFS, 11.1 vs 6.9 mo; median OS, NS</td>
</tr>
<tr>
<td>Lenvatinib for mantle-cell lymphoma</td>
<td>124 870</td>
<td>Thalidomide</td>
<td>Immune-modulatory drug (thalidomide analogue)</td>
<td>RR, 26%; median DOR, 16.6 mo</td>
</tr>
<tr>
<td>Trametinib for malignant melanoma</td>
<td>125 280</td>
<td>NA</td>
<td>First approved mek inhibitor</td>
<td>Median PFS, 4.8 vs 1.5 mo</td>
</tr>
<tr>
<td>Dabrafenib for malignant melanoma</td>
<td>109 440</td>
<td>Vemurafenib</td>
<td>BRAF inhibitor</td>
<td>Median PFS, 5.1 vs 2.7 mo; median OS, NS</td>
</tr>
<tr>
<td>Radium 223 for prostate cancer</td>
<td>82 800</td>
<td>NA</td>
<td>First approved radiotherapeutic drug</td>
<td>Median OS, 14.0 vs 11.2 mo</td>
</tr>
<tr>
<td>Erlotinib for non-small-cell lung cancer</td>
<td>82 827</td>
<td>NA</td>
<td>Block the microtubule inhibitor</td>
<td>Median PFS, 10.4 vs 5.2 mo; median OS, 12.2</td>
</tr>
<tr>
<td>Pemalidomide for multiple myeloma</td>
<td>150 408</td>
<td>Thalidomide</td>
<td>Immune-modulatory drug (thalidomide analogue)</td>
<td>RR, 29%; median DOR, 7.4 mo</td>
</tr>
<tr>
<td>Bevacizumab for colorectal cancer</td>
<td>59 422</td>
<td>NA</td>
<td>First anti-VEGF monoclonal antibody</td>
<td>Median PFS, 5.7 vs 4.0 mo; median OS, 11.2</td>
</tr>
<tr>
<td>Panitumib for chronic myeloid leukemia and Ph+ acute lymphoblastic leukemia</td>
<td>137 952</td>
<td>Imatinib</td>
<td>Bcr-abl tyrosine kinase inhibitor</td>
<td>Major cytogenetic response, 54%; median DOR, 3.2-9.5 mo</td>
</tr>
<tr>
<td>Abiraterone for prostate cancer</td>
<td>92 092</td>
<td>Ketekanazole</td>
<td>Androgen biosynthesis inhibitor</td>
<td>Median OS, 35.3 vs 30.1 mo</td>
</tr>
<tr>
<td>Cabozantinib for medullary thyroid cancer</td>
<td>118 800</td>
<td>NA</td>
<td>First multi-target VEGF and RET inhibitor</td>
<td>Median PFS, 11.2 vs 4.0 mo; median OS, NS</td>
</tr>
<tr>
<td>Omacetaxine for chronic myeloid leukemia</td>
<td>81 400</td>
<td>NA</td>
<td>Protein tyrosine kinase inhibitor</td>
<td>Major cytogenetic response, 14.3%; median DOR, 12.5 mo</td>
</tr>
<tr>
<td>Nab-paclitaxel for non-small-cell lung cancer</td>
<td>82 231</td>
<td>Paclitaxel</td>
<td>Albumin bound paclitaxel (microtubule inhibitor)</td>
<td>RR, 33% vs 25%; median OS, NS</td>
</tr>
<tr>
<td>Renografen for colorectal cancer</td>
<td>141 372</td>
<td>Sorafenib</td>
<td>Multikinase inhibitor</td>
<td>Median PFS, 21.0 vs 11.0 mo</td>
</tr>
</tbody>
</table>

*Average treatment cost $100K per year

Average PFS Improvement < 6 months

20 New Oncologic Drugs Approved Between 2009 and 2013

- Pharmacy
- Radiation
- Surgery

*This drug was approved separately for 2 indications.

---

Notes:
- PFS: Progression-Free Survival
- OS: Overall Survival
- CR: Complete Response
- PR: Partial Response
- SD: Stable Disease
- PD: Progression Disease
- DOR: Duration of Response
- OS: Overall Survival
- RR: Response Rate
- DCR: Disease Control Rate
- NED: No Evidence of Disease
Personalized Cancer Therapy

Molecular Profiling

1. Prognostic Markers
   - Markers predictive of drug sensitivity/resistance
   - Markers predictive of adverse events

2.
Drug Response is specific to Cancer (tissue) type and specific genetic variance in each tumor.

Fig. 2. Heatmap for normalized IC50 values of 75 drugs (columns) on 624 cell lines (rows). Green means the most sensitive, red means the most resistant.
Patient Derived Xenograft Models

Patient-derived xenografts (PDX) & conditionally reprogrammed cell lines

Tumorigenesis

Create reprogrammed cell lines

Transplantation into NSG mice

CL and PD Xenografts

Machine Learning In Cancer Research

- Cancer Susceptibility
- Cancer Detection and Diagnosis
- Cancer Recurrence
- Cancer Prognosis and Survival
- Cancer Classification and Clustering
- Cancer Drug Response Prediction
- Cancer Genomics Analysis
- Cancer Medical Records Analysis
- Cancer Biology

Google has developed an "Augmented Reality Microscope" that allows various deep learning algorithms to be tried on the images that it captures and for the results to be immediately seen in the microscope's field of view. Moreover, the same technology can be integrated into existing clinical microscopes. This can really help the people working on getting computers to take over a lot of the responsibilities of pathologists.
Deep Learning in Cancer ⇒ many Methods

- **AutoEncoders** – learning data representations for classification and prediction of drug response, molecular trajectories
- **VAEs and GANs** – generating data to support methods development, data augmentation and feature space algebra, drug candidate generation
- **CNNs** – type classification, drug response, outcomes prediction, drug resistance
- **RNNs** – sequence, text and molecular trajectories analysis
Machine Learning In Cancer Research

- Cancer Susceptibility
- Cancer Detection and Diagnosis
- Cancer Recurrence
- Cancer Prognosis and Survival
- Cancer Classification and Clustering
- **Cancer Drug Response Prediction**
- Cancer Genomics Analysis
- Cancer Medical Records Analysis
- Cancer Biology
Dose Response and Therapeutic Windows

We want to predict the growth rate for given drug and dose and eventually therapeutic windows.
Modeling Cancer Drug Response

\[ \mathcal{R} = f(\mathcal{T}, D_1, D_2) \]

- **Drug(s)**
  - descriptors
  - fingerprints
  - structures
  - SMILES
  - dose

- **Tumor**
  - IC50
  - GI50
  - % growth
  - Z-score

- **Response**
  - gene expression levels
  - SNPs
  - protein abundance
  - microRNA
  - methylation
Top 29 Single Drug ML Models

Many Methods

(No Free Lunch)
Ensemble ML Model for Predicting Narciclasine Response

N=741
AUC ~0.88
5-Fold Cross Validation Accuracy ~83.6%

Narciclasine is a toxic alkaloid found in various Amaryllidaceae species. Wikipedia
Formula: \( \text{C}_{14}\text{H}_{13}\text{NO}_7 \)
IUPAC ID: \((2S\{2a,3b,4\theta,4a\beta\})-3,4,4a,5\text{-tetrahydro}-2,3,4,7\text{-tetrahydroxy}-\{(1,3\text{-dioxol}(4,5,6)\text{phenanthridin}(6\text{-H})-\text{one})\)
PubChem CID: 72376
Selection Summary

- Maximizes F1 Score
- NOT(RS) + RS

F1 Score: 0.7298
True Positive Rate (Sensitivity): 0.7278
False Positive Rate (Fallout): 0.1165
True Negative Rate (Specificity): 0.8835
Positive Predictive Value (Precision): 0.7318
Negative Predictive Value: 0.8814
Accuracy: 0.8361
Matthews Correlation Coefficient: 0.6122

Actual:
- Predicted: 364 (TN) 48 (FP) 412
- + 49 (FN) 131 (TP) 180

ROC Curve
Data Source: Cross Validation
Threshold (0-1): 0.3598

Prediction Distribution
Density

Narciscasine

Narciscasine is a toxic alkaloid found in various Amaryllidaceae species. Wikipedia

Formula: C_{44}H_{53}NO_{7}
IUPAC ID: (2S*(2a,3a,4,4a8))\cdot3,4,4a,5-Tetrahydro-2,3,4,7-tetrahydroxy-(1,3)oxol(o(4,5))-phenanthridine-6(2H)-one
PubChem CID: 72378
S100 Calcium Binding Protein A13

Involved in the regulation of a number of cellular processes such as cell cycle progression and differentiation.
S100 Calcium Binding Protein A13

Higher Expression Levels $\implies$ Lowered Response to Drug

Response

Expression Level
Renal Cancer

1.97e-4 (Prognostic, unfavourable)
<table>
<thead>
<tr>
<th>Q</th>
<th>R</th>
<th>S</th>
<th>T</th>
<th>U</th>
<th>V</th>
<th>W</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>AA</th>
<th>AB</th>
<th>AC</th>
<th>AD</th>
<th>AE</th>
<th>AF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paclitaxel</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>1.00</td>
<td>0.86</td>
<td>0.71</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.86</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>1.00</td>
<td>0.86</td>
<td>0.71</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
</tbody>
</table>
We a single model that can train on data from many cancer samples, many drugs and can predict drug response across a wide range of tumors and drug combinations.
Why deep learning

Performance vs. Amount of data

Deep learning

Other learning algorithms
Model Accuracy as a Function of Training Set Size

MINIST (10 digit) Accuracy for Exponentially Increasing Dataset Sizes

MNIST Dataset
10 digits
6,000 examples of each
best models are 99%+

Note: Log Scale
How much data do we need?

• Three heuristics that are sometimes used
  
  • X times the number of classes in the data \( \{X \sim 1000, X \sim 10,000\} \)
    
    \((\text{Drugs } \times \text{Cancer Samples } \times \text{Dose } \times \text{Response categories}) \sim 100M - \sim 1B\)
  
  • X times the number of features \( \{X \sim 100, X \sim 1000\} \)
    
    \((10,000 \times 100) \sim 1M - \sim 10M\)
  
  • X times the number of model parameters \( \{X \sim 10, X \sim 100\} \)
    
    \((10M \times 10) \sim 100M \text{ to } (100M \times 100) \sim 10B\)

• 1M to 10B training examples

• Current training sets are in the low end of this range
<table>
<thead>
<tr>
<th>Data Source</th>
<th># Tumor Samples</th>
<th># Drugs</th>
<th># Dose Response Samples</th>
<th>Treatment Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCI-ALMANAC</td>
<td>60</td>
<td>104</td>
<td>3,686,475</td>
<td>Drug pair</td>
</tr>
<tr>
<td>CCLE</td>
<td>504</td>
<td>24</td>
<td>93,251</td>
<td>Single drug</td>
</tr>
<tr>
<td>CTRPv2</td>
<td>887</td>
<td>544</td>
<td>6,171,005</td>
<td>Single drug</td>
</tr>
<tr>
<td>gCSI</td>
<td>409</td>
<td>16</td>
<td>58,094</td>
<td>Single drug</td>
</tr>
<tr>
<td>GDSC</td>
<td>1,075</td>
<td>249</td>
<td>1,894,212</td>
<td>Single drug</td>
</tr>
<tr>
<td>NCI</td>
<td>60</td>
<td>52,671</td>
<td>18,862,308</td>
<td>Single drug</td>
</tr>
<tr>
<td>GDC</td>
<td>11,081</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>NCI-PDM</td>
<td>1,198</td>
<td>12</td>
<td>518*</td>
<td>Single and paired drugs</td>
</tr>
</tbody>
</table>

* PDM drug response were measured differently from cell line dose response data.
Deep Learning Model for Drug Pair Response

Fig. 2. **Neural network architecture.** The orange square boxes, from bottom to top, represent input features, encoded features, and output growth values. Feature models are denoted by round shaded boxes: green for molecular features and blue for drug features. There are multiple types of molecular features that are fed into submodels for gene expression, proteome, and microRNA. The descriptors for the two drugs share the same descriptor model. All encoded features are then concatenated to form input for the top fully connected layers. Most connecting layers are linked by optional residual skip connections if their dimensions match.
Drug “Synergy”

Combination index $CI = \frac{x_1}{X_1 + x_2/X_2}$

- **Synergy**: $CI = 1$
- **Additive**: $CI = 1$
- **Antagonism**: $CI > 1$

Synergy
Independent
Antagonism

$$Y_{ab,O} \begin{cases} > Y_{ab,P} \\ = Y_{ab,P} \\ < Y_{ab,P} \end{cases}$$

Graph showing the relationship between median tumor volume and days post implantation for different treatment groups.
Deep Learning Model for Drug Pair Response

Table 1 Cross validation results from feature combination experiments

<table>
<thead>
<tr>
<th>Molecular Features</th>
<th>Drug Features</th>
<th>MSE</th>
<th>MAE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>baseline</td>
<td>0.5253</td>
<td>0.5709</td>
<td>-1.001</td>
</tr>
<tr>
<td>one-hot encoding</td>
<td>one-hot encoding</td>
<td>0.2448</td>
<td>0.3997</td>
<td>0.1269</td>
</tr>
<tr>
<td>gene expression</td>
<td>one-hot encoding</td>
<td>0.2447</td>
<td>0.3999</td>
<td>0.1272</td>
</tr>
<tr>
<td>gene expression</td>
<td>500-dimensional noise</td>
<td>0.2450</td>
<td>0.4008</td>
<td>0.1271</td>
</tr>
<tr>
<td>one-hot encoding</td>
<td>Dragon7 descriptors</td>
<td>0.0292</td>
<td>0.1086</td>
<td>0.8892</td>
</tr>
<tr>
<td>proteome</td>
<td>Dragon7 descriptors</td>
<td>0.0303</td>
<td>0.1117</td>
<td>0.8844</td>
</tr>
<tr>
<td>microRNA</td>
<td>Dragon7 descriptors</td>
<td>0.0275</td>
<td>0.1050</td>
<td>0.8952</td>
</tr>
<tr>
<td>gene expression</td>
<td>Dragon7 descriptors</td>
<td>0.0180</td>
<td>0.0906</td>
<td>0.9364</td>
</tr>
<tr>
<td>gene expression, microRNA, proteome</td>
<td>Dragon7 descriptors</td>
<td><strong>0.0158</strong></td>
<td><strong>0.0833</strong></td>
<td><strong>0.9440</strong></td>
</tr>
</tbody>
</table>

DNN Model explains 94% of the variance
Do ML models transfer across studies? Do ML models transfer across “bio” model Types?

Transfer Learning and Model Transfer

“transfer learning” is using training data from another (possibly related) area to accelerate training and improve generalizability, generally requires additional training in the target domain.

“model transfer” is using models trained in one area to predict in another without tuning in the target domain.

Until we have sufficient data from PDXs to “tune” models trained with CL we are in strong “model transfer” regime.
Cross Study Validation Targets (Cell Lines)
Batch Effects Removal
Cross Study Validation – Models are trained on one study and predict on the other studies not used in training (strong model transfer)

<table>
<thead>
<tr>
<th>Training set</th>
<th>NCI60</th>
<th>CTRP</th>
<th>GDSC</th>
<th>CCLE</th>
<th>gCSI</th>
<th># Drugs</th>
<th># Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCI60</td>
<td>R² = 0.74, MAE = 21</td>
<td>R² = 0.32, MAE = 36</td>
<td>R² = 0.20, MAE = 38</td>
<td>R² = 0.38, MAE = 37</td>
<td>R² = 0.42, MAE = 35</td>
<td>1006</td>
<td>59</td>
</tr>
<tr>
<td>CTRP</td>
<td>R² = 0.42, MAE = 31</td>
<td>R² = 0.67, MAE = 23</td>
<td>R² = 0.25, MAE = 35</td>
<td>R² = 0.56, MAE = 30</td>
<td>R² = 0.54, MAE = 31</td>
<td>495</td>
<td>812</td>
</tr>
<tr>
<td>GDSC</td>
<td>R² = 0.30, MAE = 36</td>
<td>R² = 0.19, MAE = 39</td>
<td>R² = 0.55, MAE = 27</td>
<td>R² = 0.41, MAE = 36</td>
<td>R² = 0.58, MAE = 29</td>
<td>239</td>
<td>672</td>
</tr>
<tr>
<td>CCLE</td>
<td>R² = 0.05, MAE = 43</td>
<td>R² = -0.07, MAE = 45</td>
<td>R² = -0.21, MAE = 46</td>
<td>R² = 0.60, MAE = 29</td>
<td>R² = 0.20, MAE = 39</td>
<td>24</td>
<td>474</td>
</tr>
<tr>
<td>gCSI</td>
<td>R² = 0.19, MAE = 41</td>
<td>R² = -0.02, MAE = 48</td>
<td>R² = -0.03, MAE = 45</td>
<td>R² = 0.11, MAE = 46</td>
<td>R² = 0.48, MAE = 37</td>
<td>16</td>
<td>357</td>
</tr>
<tr>
<td>Combined</td>
<td>R² = 0.66, MAE = 23</td>
<td>R² = 0.60, MAE = 26</td>
<td>R² = 0.44, MAE = 31</td>
<td>R² = 0.59, MAE = 30</td>
<td>R² = 0.61, MAE = 29</td>
<td>1780</td>
<td>2374</td>
</tr>
</tbody>
</table>
Can we create a unified deep learning model to solve tasks across multiple domains?

We demonstrate, for the first time, that a single deep learning model can jointly learn a number of large-scale tasks from multiple domains. The key to success comes from designing a multi-modal architecture in which as many parameters as possible are shared and from using computational blocks from different domains together. We believe that this treads a path towards interesting future work on more general deep learning architectures, especially since our model shows transfer learning from tasks with a large amount of available data to ones where the data is limited.
Figure 7. An improved DNN with drug target input and additional tumor properties. We extended the Combo drug response model delivered in Milestone 4 by (1) unifying single and paired drug response, (2) adding drug concentration, (3) adding drug target descriptions to the input, and (4) adding four optional prediction targets (tissue category, tumor site, cancer type and gene expression autoencoder) in a multitasking framework. The purple boxes delineate the changes in the new Uno models.
Baseline Random Forest Cross-Study Run
Best out of study $R^2 = 0.45$

Table 3. Baseline cross study validation results with Random Forest

<table>
<thead>
<tr>
<th>Training set</th>
<th>Testing set</th>
<th>NCI60</th>
<th>CTRP</th>
<th>GDSC</th>
<th>CCLE</th>
<th>gCSI</th>
<th># Drugs</th>
<th># Cells</th>
<th># Pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCI60</td>
<td>NCI60</td>
<td>$R^2 = 0.45$</td>
<td>$R^2 = 0.23$</td>
<td>$R^2 = 0.15$</td>
<td>$R^2 = 0.29$</td>
<td>$R^2 = 0.14$</td>
<td>1006</td>
<td>59</td>
<td>59364</td>
</tr>
<tr>
<td>CTRP</td>
<td>CTRP</td>
<td>$R^2 = 0.41$</td>
<td>$R^2 = 0.30$</td>
<td>$R^2 = 0.15$</td>
<td>$R^2 = 0.45$</td>
<td>$R^2 = 0.17$</td>
<td>495</td>
<td>812</td>
<td>401940</td>
</tr>
<tr>
<td>GDSC</td>
<td>GDSC</td>
<td>$R^2 = 0.33$</td>
<td>$R^2 = 0.14$</td>
<td>$R^2 = 0.13$</td>
<td>$R^2 = 0.17$</td>
<td>$R^2 = 0.08$</td>
<td>239</td>
<td>672</td>
<td>160608</td>
</tr>
<tr>
<td>CCLE</td>
<td>CCLE</td>
<td>$R^2 = 0.12$</td>
<td>$R^2 = -0.03$</td>
<td>$R^2 = -0.11$</td>
<td>$R^2 = 0.17$</td>
<td>$R^2 = 0.32$</td>
<td>24</td>
<td>474</td>
<td>11376</td>
</tr>
<tr>
<td>gCSI</td>
<td>gCSI</td>
<td>$R^2 = -0.38$</td>
<td>$R^2 = -0.51$</td>
<td>$R^2 = -0.59$</td>
<td>$R^2 = -0.09$</td>
<td>$R^2 = 0.25$</td>
<td>16</td>
<td>357</td>
<td>5712</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Training set</th>
<th>Testing set</th>
<th>NCI60</th>
<th>CTRP</th>
<th>GDSC</th>
<th>CCLE</th>
<th>gCSI</th>
<th># Drugs</th>
<th># Cells</th>
<th># Pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCI60</td>
<td>NCI60</td>
<td>$MAE = 30.4$</td>
<td>$MAE = 34.6$</td>
<td>$MAE = 37.3$</td>
<td>$MAE = 36.4$</td>
<td>$MAE = 54.0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTRP</td>
<td>CTRP</td>
<td>$MAE = 31.7$</td>
<td>$MAE = 35.0$</td>
<td>$MAE = 37.4$</td>
<td>$MAE = 36.4$</td>
<td>$MAE = 54.0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDSC</td>
<td>GDSC</td>
<td>$MAE = 36.0$</td>
<td>$MAE = 41.5$</td>
<td>$MAE = 40.4$</td>
<td>$MAE = 42.4$</td>
<td>$MAE = 43.0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCLE</td>
<td>CCLE</td>
<td>$MAE = 42.6$</td>
<td>$MAE = 48.9$</td>
<td>$MAE = 47.1$</td>
<td>$MAE = 42.4$</td>
<td>$MAE = 38.5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gCSI</td>
<td>gCSI</td>
<td>$MAE = 55.0$</td>
<td>$MAE = 59.0$</td>
<td>$MAE = 58.7$</td>
<td>$MAE = 48.6$</td>
<td>$MAE = 39.9$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# UnoMT Multitask Deep Learning Cross-Study

Best out of Study $R^2 = 0.61$

## Table 6. Best cross study validation results with a 3-task UnoMT

<table>
<thead>
<tr>
<th>Training set</th>
<th>NCI60</th>
<th>CTRP</th>
<th>GDSC</th>
<th>CCLE</th>
<th>gCSI</th>
<th>aux tasks</th>
<th>N/T Cat Acc</th>
<th>Site Acc</th>
<th>Type Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCI60</td>
<td>R2 = 0.81, MAE = 17.1</td>
<td>R2 = 0.38, MAE = 32.2</td>
<td>R2 = 0.24, MAE = 35.3</td>
<td>R2 = 0.48, MAE = 33.4</td>
<td>R2 = 0.46, MAE = 33.4</td>
<td>N/T: 99.43%, Site: 96.75%, Type: 96.97%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTRP</td>
<td>R2 = 0.44, MAE = 29.8</td>
<td>R2 = 0.68, MAE = 22.7</td>
<td>R2 = 0.23, MAE = 34.4</td>
<td>R2 = 0.61, MAE = 28.3</td>
<td>R2 = 0.60, MAE = 28.5</td>
<td>N/T: 99.56%, Site: 96.62%, Type: 96.58%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDSC</td>
<td>R2 = 0.32, MAE = 34.0</td>
<td>R2 = 0.25, MAE = 36.7</td>
<td>R2 = 0.53, MAE = 27.2</td>
<td>R2 = 0.50, MAE = 32.6</td>
<td>R2 = 0.60, MAE = 29.2</td>
<td>N/T: 99.43%, Site: 96.93%, Type: 96.97%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCLE</td>
<td>R2 = 0.27, MAE = 36.9</td>
<td>R2 = 0.20, MAE = 39.2</td>
<td>R2 = 0.11, MAE = 38.9</td>
<td>R2 = 0.68, MAE = 25.4</td>
<td>R2 = 0.39, MAE = 34.2</td>
<td>N/T: 99.12%, Site: 96.36%, Type: 96.36%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gCSI</td>
<td>R2 = 0.00, MAE = 44.9</td>
<td>R2 = 0.11, MAE = 43.1</td>
<td>R2 = 0.05, MAE = 42.8</td>
<td>R2 = 0.33, MAE = 40.6</td>
<td>R2 = 0.80, MAE = 19.2</td>
<td>N/T: 99.43%, Site: 96.84%, Type: 96.62%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Machine Learning Models with UQ

[High-]Throughput Experiments

Interesting Biology

Model Uncertainty

Training Data

Additional Training Data
P1 Challenge Problem Workflow(s) Specification

**Data Preparation**
- Batch Normalization
- Data Augmentation
- Outlier Removal
- Scaling/Quantization
- Concordance Processing

**Model Discovery**
- Residual Networks
- Convolution
- Multitask Networks
- Population Based HPO

**Training**
- Ensembles
- Domain Adaptation
- Cross-validation
- Transfer Learning
- UQ
- Factorial Design
- Learning Curves

**Inference**
- Source – Target Pairs
- Drug Combinations
- Confidence Scoring

**Outputs**
- Accuracy / K-rank / $R^2$
- Feature importance
- Performance Analysis
P1 Challenge Problem Workflow(s) Specification

Data Preparation
- Batch Normalization
- Data Augmentation
- Outlier Removal
- Scaling/Quantization
- Concordance Processing

Model Discovery
- Multitask Networks
- Population Based HPO
- $10^5 - 10^6$ units of work

Training
- Ensembles
- Domain Adaptation
- Cross-validation
- $10^5 - 10^6$ units of work
- Factorial Design
- Learning Curves

Inference
- $10^6 - 10^8$ units of work
- Confidence Scoring

Outputs
- Accuracy / K-rank / $R^2$
- Feature importance
- Performance Analysis
CANDLE Challenge Problem Statement

Enable the most challenging deep learning problems in Cancer research to be pursued on the most capable supercomputers in the DOE
ECP-CANDLE: CANcer Distributed Learning Environment

CANDLE Approach

Develop an exscale deep learning environment for cancer
Build on open source deep learning frameworks
Optimize for CORAL and Exascale platforms
Support all three pilot project needs for deep dearning
Collaborate with DOE computing centers, HPC vendors and ECP co-design and software technology projects
CANDLE Components

• **CANDLE Python Library** – make it easy to run on DOE Big Machines, scale for HPO, UQ, Ensembles, Data Management, Logging, Analysis

• **CANDLE Benchmarks** – exemplar codes/models and data representing the three primary challenge problems

• **Runtime Software** – Supervisor, Reporters, Data Management, Run Data Base

• **Tutorials** – Well documented examples for engaging the community

• **Contributed Codes** – Examples outside of Cancer, including Climate Research, Materials Science, Imaging, Brain Injury

• **Frameworks** – Leverage of Tensorflow, Keras, Horovod, LBANN, etc.

• **LL Libraries** – CuDNN, MKL, etc. (tuned to DOE machines)
<table>
<thead>
<tr>
<th></th>
<th>Languages</th>
<th>Tutorials and training materials</th>
<th>CNN modeling capability</th>
<th>RNN modeling capability</th>
<th>Architecture: easy-to-use and modular front end</th>
<th>Speed</th>
<th>Multiple GPU support</th>
<th>Keras compatible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theano</td>
<td>Python, C++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Tensor-Flow</td>
<td>Python</td>
<td>+++</td>
<td>+++</td>
<td>++</td>
<td>+++</td>
<td>++</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>Torch</td>
<td>Lua, Python (new)</td>
<td>+</td>
<td>+++</td>
<td>++</td>
<td>++</td>
<td>+++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Caffe</td>
<td>C++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MXNet</td>
<td>R, Python, Julia, Scala</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>+++</td>
<td>+</td>
</tr>
<tr>
<td>Neon</td>
<td>Python</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CNTK</td>
<td>C++</td>
<td>+</td>
<td>+</td>
<td>+++</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Candle Functional Targets

- Enable high productivity for deep learning centric workflows
- Support Key DL frameworks on DOE supercomputers (Keras, TF, Mxnet, CNTK)
- Support multiple paths to concurrency (Ensembles, Data and Model Parallel)
- Manage training data, model search, scoring, optimization, production training and inference (End-to-End Workflow)
- CANDLE runtime/supervisor (interface with batch schedulers)
- CANDLE Python library for improving model development (UQ, HPO, CV, MV)
- Well documented open examples and tutorials on Github
- Leverage as much open source as possible (build only what we need to add to existing frameworks)
<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Type</th>
<th>Data</th>
<th>ID</th>
<th>OD</th>
<th>Sample Size</th>
<th>Size of Network</th>
<th>Additional (activation, layer types, etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. P1: B1 Autoencoder</td>
<td>MLP</td>
<td>RNA-Seq</td>
<td>$10^5$</td>
<td>$10^5$</td>
<td>15K</td>
<td>5 layers</td>
<td>Log2 (x+1) (\rightarrow) [0,1] KPRM-UQ</td>
</tr>
<tr>
<td>2. P1: B2 Classifier</td>
<td>MLP</td>
<td>SNP</td>
<td>$10^5$</td>
<td>$10^5$</td>
<td>40</td>
<td>5 layers</td>
<td>Training Set Balance issues</td>
</tr>
<tr>
<td>3. P1: B3 Regression</td>
<td>MLP+LCN</td>
<td>expression; drug desc</td>
<td>$10^5$</td>
<td>$10^5$</td>
<td>3M</td>
<td>8 layers</td>
<td>Drug Response [-100, 100]</td>
</tr>
<tr>
<td>4. P2: B1 Autoencoder</td>
<td>MLP</td>
<td>MD K-RAS</td>
<td>$10^5$</td>
<td>$10^5$</td>
<td>$10^6$-$10^8$</td>
<td>5-8 layers</td>
<td>State Compression</td>
</tr>
<tr>
<td>5. P2: B2 RNN-LSTM</td>
<td>RNN-LSTM</td>
<td>MD K-RAS</td>
<td>$10^5$</td>
<td>$10^5$</td>
<td>$10^6$</td>
<td>4 layers</td>
<td>State to Action</td>
</tr>
<tr>
<td>6. P3: B1 RNN-LSTM</td>
<td>RNN-LSTM</td>
<td>Path reports</td>
<td>$10^3$</td>
<td>$10^3$</td>
<td>$5\times10^5$</td>
<td>1-2 layers</td>
<td>Dictionary 12K +30K</td>
</tr>
<tr>
<td>7. P3: B2 Classification</td>
<td>CNN</td>
<td>Path reports</td>
<td>$10^4$</td>
<td>$10^4$</td>
<td>$10^5$</td>
<td>5 layers</td>
<td>Biomarkers</td>
</tr>
</tbody>
</table>

**Benchmark Owners:**
- P1: Fangfang Xia (ANL)
- P2: Brian Van Essen (LLNL)
- P3: Arvind Ramanathan (ORNL)

**Drug Response**

**RAS Pathways**

**Patient Trajectories**
BFP16 Probably as Good as FP32 for Training
<table>
<thead>
<tr>
<th>Operation</th>
<th>Size</th>
<th>Sum</th>
<th>Fraction</th>
<th>Time</th>
<th>Sum-Time</th>
<th>Fract-Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>MatMul</td>
<td>121.67MB</td>
<td>(100.00%, 33.33%)</td>
<td>115.52ms</td>
<td>(100.00%, 40.21%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assign</td>
<td>0B</td>
<td>(0.00%, 0.00%)</td>
<td>73.65ms</td>
<td>(59.79%, 25.64%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mul</td>
<td>121.02MB</td>
<td>(66.67%, 33.15%)</td>
<td>61.46ms</td>
<td>(34.15%, 21.39%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add</td>
<td>120.35MB</td>
<td>(33.52%, 32.97%)</td>
<td>22.31ms</td>
<td>(12.75%, 7.77%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sub</td>
<td>0B</td>
<td>(0.00%, 0.00%)</td>
<td>2.22ms</td>
<td>(4.98%, 0.77%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RandomUniform</td>
<td>660.00KB</td>
<td>(0.54%, 0.18%)</td>
<td>1.75ms</td>
<td>(4.21%, 0.61%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RealDiv</td>
<td>660.40KB</td>
<td>(0.36%, 0.18%)</td>
<td>1.75ms</td>
<td>(3.60%, 0.61%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BiasAdd</td>
<td>0B</td>
<td>(0.00%, 0.00%)</td>
<td>1.19ms</td>
<td>(3.00%, 0.41%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ReluGrad</td>
<td>0B</td>
<td>(0.00%, 0.00%)</td>
<td>1.11ms</td>
<td>(2.58%, 0.39%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BiasAddGrad</td>
<td>6.60KB</td>
<td>(0.18%, 0.00%)</td>
<td>1.00ms</td>
<td>(2.20%, 0.35%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VariableV2</td>
<td>0B</td>
<td>(0.00%, 0.00%)</td>
<td>868us</td>
<td>(1.85%, 0.30%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fill</td>
<td>660.00KB</td>
<td>(0.18%, 0.00%)</td>
<td>849us</td>
<td>(1.55%, 0.30%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floor</td>
<td>0B</td>
<td>(0.00%, 0.00%)</td>
<td>848us</td>
<td>(1.25%, 0.30%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P1B3 Operation Profile
## P1B3 Matrix Sizes and Times for One Pass on x86

<table>
<thead>
<tr>
<th>Rows</th>
<th>Columns</th>
<th>Parameters</th>
<th>Execution Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>0:100x100,</td>
<td>1:100x50</td>
<td>(run*2</td>
<td>defined*2)</td>
</tr>
<tr>
<td>0:-1x29532,</td>
<td>1:100x1000</td>
<td>(run*1</td>
<td>defined*1)</td>
</tr>
<tr>
<td>0:-1x29532,</td>
<td>1:29532x1000</td>
<td>(run*1</td>
<td>defined*1)</td>
</tr>
<tr>
<td>0:100x1,</td>
<td>1:50x1</td>
<td>(run*1</td>
<td>defined*1)</td>
</tr>
<tr>
<td>0:100x100,</td>
<td>1:500x100</td>
<td>(run*1</td>
<td>defined*1)</td>
</tr>
<tr>
<td>0:100x1000,</td>
<td>1:1000x500</td>
<td>(run*1</td>
<td>defined*1)</td>
</tr>
<tr>
<td>0:100x1000,</td>
<td>1:100x500</td>
<td>(run*1</td>
<td>defined*1)</td>
</tr>
<tr>
<td>0:100x50,</td>
<td>1:100x1</td>
<td>(run*1</td>
<td>defined*1)</td>
</tr>
<tr>
<td>0:100x50,</td>
<td>1:100x50</td>
<td>(run*1</td>
<td>defined*1)</td>
</tr>
<tr>
<td>0:100x50,</td>
<td>1:50x1</td>
<td>(run*1</td>
<td>defined*1)</td>
</tr>
<tr>
<td>0:100x500,</td>
<td>1:1000x500</td>
<td>(run*1</td>
<td>defined*1)</td>
</tr>
<tr>
<td>0:100x500,</td>
<td>1:100x100</td>
<td>(run*1</td>
<td>defined*1)</td>
</tr>
<tr>
<td>0:100x500,</td>
<td>1:500x100</td>
<td>(run*1</td>
<td>defined*1)</td>
</tr>
<tr>
<td>0:100x1000,</td>
<td>1:29532x1000</td>
<td>(run*0</td>
<td>defined*1)</td>
</tr>
</tbody>
</table>
CANDLE System Architecture

CANDLE Specifications
- Benchmark Spec
- Hyperparameter Spec
- Hardware Spec

CANDLE Supervisor
- Hyperparameter Optimization Frameworks
  - Hyperopt, mlrMBO, Spearmint
- Workflow Manager
  (Swift-T EMEWS)

ML/DL Benchmarks
- Pilot 1
- Pilot 2
- Pilot 3

CANDLE Database
- Metadata Store
  - Benchmarks
  - Datasets
  - Models
  - Experiments
  - Runs
- Model Store
  - Model Descriptions
  - Model Weights
- Data API

Integrator Website

Hardware Resources
- ALCF
  - Theta, Cooley
- NERSC
  - Cori
- OLCF
  - Titan, SummitDev

Hyperparameter Optimization Frameworks
- Hyperopt
- mlrMBO
- Spearmint

Benchmarks
- Datasets
- Models
- Experiments
- Runs

CANDLE Specifications

ALCF
  - Theta, Cooley

NERSC
  - Cori

OLCF
  - Titan, SummitDev

Hardware Resources
GitHub and FTP

• ECP-CANDLE GitHub Organization:
  • https://github.com/ECP-CANDLE

• ECP-CANDLE FTP Site:
  • The FTP site hosts all the public datasets for the benchmarks
Basic Take Away Points

• Cancer changes cell behavior. We can assay gene expression, SNPs, protein abundance, etc. to characterize these changes. Assays are averaged over the cells in a sample.

• Molecular assay data can be used to predict properties of patient tumors: Cancer Type, Cancer Site, Normal vs Tumor etc. (These predictors are quite accurate when trained on large-scale (GDC) data 98%-99% accurate)

• Model systems (Cell Lines, Organoids, Xenografts) resemble patient tumors from gene expression profiles and are assayed in the same way.

• Drug responses of “Biological” Models are similar to patient response in some cases (how similar is open question)

• Machine Learning Models can predict drug response of “Biological” Models

• ML can be used to generate and screen drugs for development

• ML eventually can be used to select drugs and drug combinations for a patients
Many thanks to DOE, NSF, NIH, DOD, ANL, UC, Moore Foundation, Sloan Foundation, Apple, Microsoft, Cray, Intel and IBM for supporting my research group over the years
Three Approaches to Uncertainty Quantification

• Train on distributions and predict distributions

• Bootstrap with ensembles during training

• Dropout during inference as a Bayesian approximation (Yarin Gal, University of Cambridge)
Figure 2. An example of dose response data from multiple studies. The figure adapted from PharmacoDB [2] shows the fitted dose response curves of the SU-DHL-8 lymphoma cell line treated with paclitaxel. Experimental measurements from multiple sources are not in complete agreement.
Intuition behind UQ (Gaussian Process Models)
Bootstrapping UQ in Deep Neural Networks

(b) Gaussian process posterior    (c) Bootstrapped neural nets

Gerhard Paass, Assessing and Improving Neural Network Predictions by the Bootstrap Algorithm, NIPS, 659
Combined Synergy and Uncertainty Map
• Student-Teacher Approach to use Mechanistic Models as Oracles for a subset of the conditions (i.e. where we have very good predictive skill for a sub-problem)
• Integration of pathway Information as constraints/hints to the machine learning models either explicitly or implicitly
• To fill-in gaps in the mechanistic models with machine learned functions
Figure 1: Dropout Neural Net Model. **Left:** A standard neural net with 2 hidden layers. **Right:** An example of a thinned net produced by applying dropout to the network on the left. Crossed units have been dropped.
Dropout vs Bootstrap

Predicted Synergy for Bootstrap N=10 (Top 50K Synergy Samples)

Predicted Synergy for Dropout N=100 (Top 50K Synergy Samples)
Order coherence and calibration

Highly confident predictions (small bootstrap std) have high accuracy (with high confidence the predictions are in a small interval around the true value).
Three Approaches to Uncertainty Quantification

- Train on distributions and predict distributions
- Bootstrap with ensembles during training
- Dropout during inference as a Bayesian approximation

(Yarin Gal, University of Cambridge)
Figure 2. An example of dose response data from multiple studies. The figure adapted from PharmacoDB [2] shows the fitted dose response curves of the SU-DHL-8 lymphoma cell line treated with paclitaxel. Experimental measurements from multiple sources are not in complete agreement.
Intuition behind UQ (Gaussian Process Models)
Bootstrapping UQ in Deep Neural Networks

(b) Gaussian process posterior  (c) Bootstrapped neural nets

Gerhard Paass, Assessing and Improving Neural Network Predictions by the Bootstrap Algorithm, NIPS, 659
Dropout!

Srivastava, Hinton, Krizhevsky, Sutskever and Salakhutdinov

Figure 1: Dropout Neural Net Model. Left: A standard neural net with 2 hidden layers. Right: An example of a thinned net produced by applying dropout to the network on the left. Crossed units have been dropped.
Dropout vs Bootstrap

Predicted Synergy for Bootstrap N=10 (Top 50K Synergy Samples)

Predicted Synergy for Dropout N=100 (Top 50K Synergy Samples)
Order coherence and calibration

Highly confident predictions (small bootstrap std) have high accuracy (with high confidence the predictions are in a small interval around the true value).