Deep Bayesian Neural Networks for improved Treatment Assignment in Precision Oncology. A Contextual Bandit Problem.

Niklas Rindtorff[1,2,6,7], Ming Yu Lu[1,5,#], Nisarg Patel[1,2,4,#], HuaHua Zheng[1,3,#], Kun-Hsing Yu[1], Alexander D'Amour[8]

[1]Harvard Medical School, Department of Biomedical Informatics, Boston, MA
[2]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[3]Harvard T.H. Chan School of Public Health, Department of Biostatistics, Boston, MA
[4]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA
[5]Laboratory for Computational Physiology, Massachusetts Institute of Technology, MA, USA
[6]German Cancer Research Center (DKFZ), Division Signaling and Functional Genomics, Heidelberg, Germany
[7]Heidelberg University, Medical Faculty Heidelberg, MD/PhD Program
[8]Google Brain, Cambridge, MA
[#]In alphabetical order

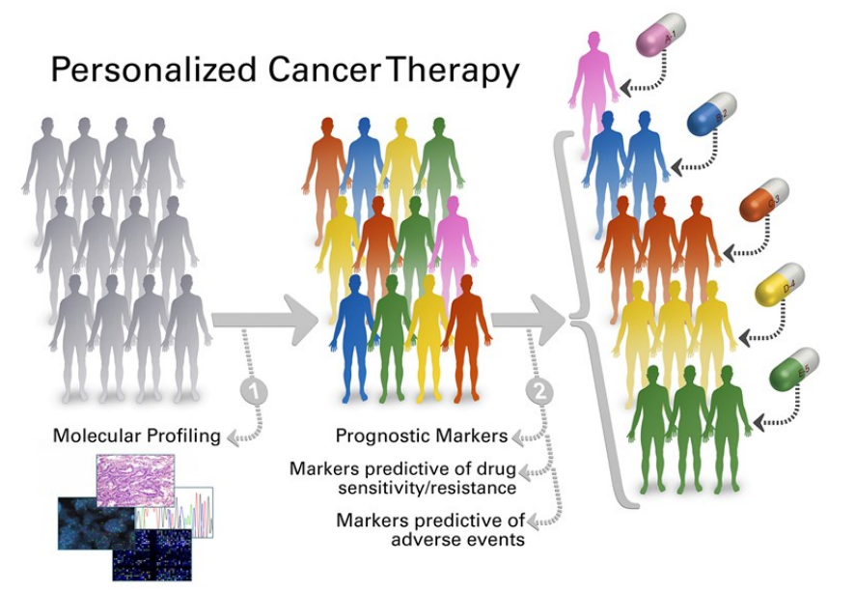
Please share!

Problem - Contextual Bandit in Precision Oncology programs

The goal of precision medicine is providing the right treatment to the right patient at the right time. Despite a number of successes, assigning patients to adequate treatments remains a challenge until today. The current best practice in precision oncology is to base the treatment decision on published and frequently used therapeutic protocols that consider the patient's clinical characteristics and cancer biomarkers. For example, based on the status of a single mutation, such as a BRAF V600E, a treatment decision can be made.

Current therapeutic protocols in precision oncology evaluate one biomarker and one targeted therapeutic at a time. This limits the ability to make high-confidence clinical decisions in a real world scenario with a large number of biomarkers to measure and potential treatments to choose from. The current limitations include:

(I) high selectiveness
(II) high unsystematic compassionate use
(III) limited continuous development
(IV) limited predictive availability

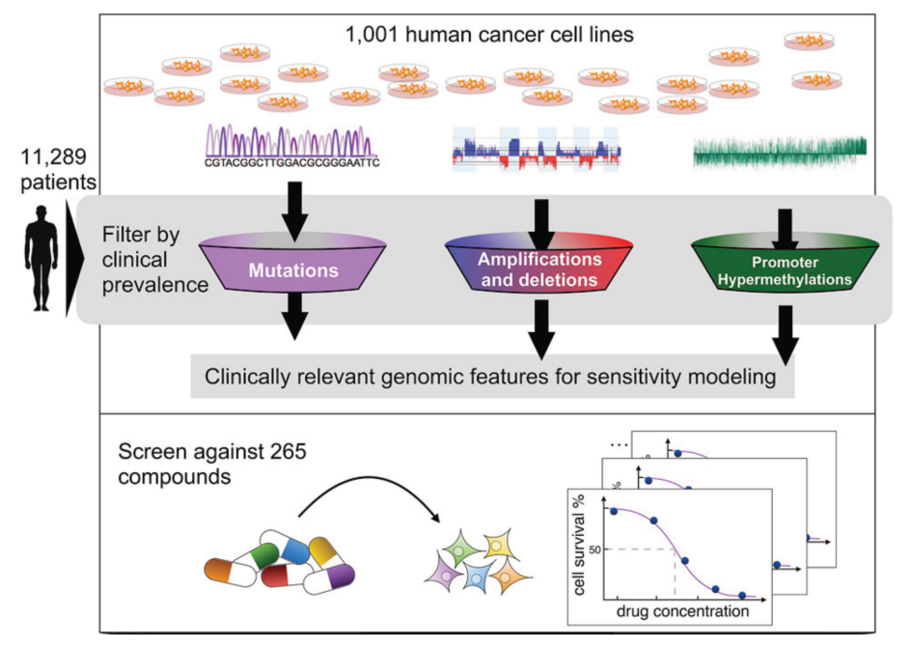


Personalized Cancer Therapy
Molecular Profiling
Prognostic Markers
Markers predictive of drug sensitivity/resistance
Markers predictive of adverse events

2018 The University of Texas MD Anderson Cancer Center

We hypothesized that, if framed as a contextual bandit problem, neural network based agents can outperform current clinical therapeutic assignment mechanisms. Potentially, this could allow a less selective, more systematic, continuously evolving practice of precision oncology, which balances exploration and exploitation of treatment strategies.

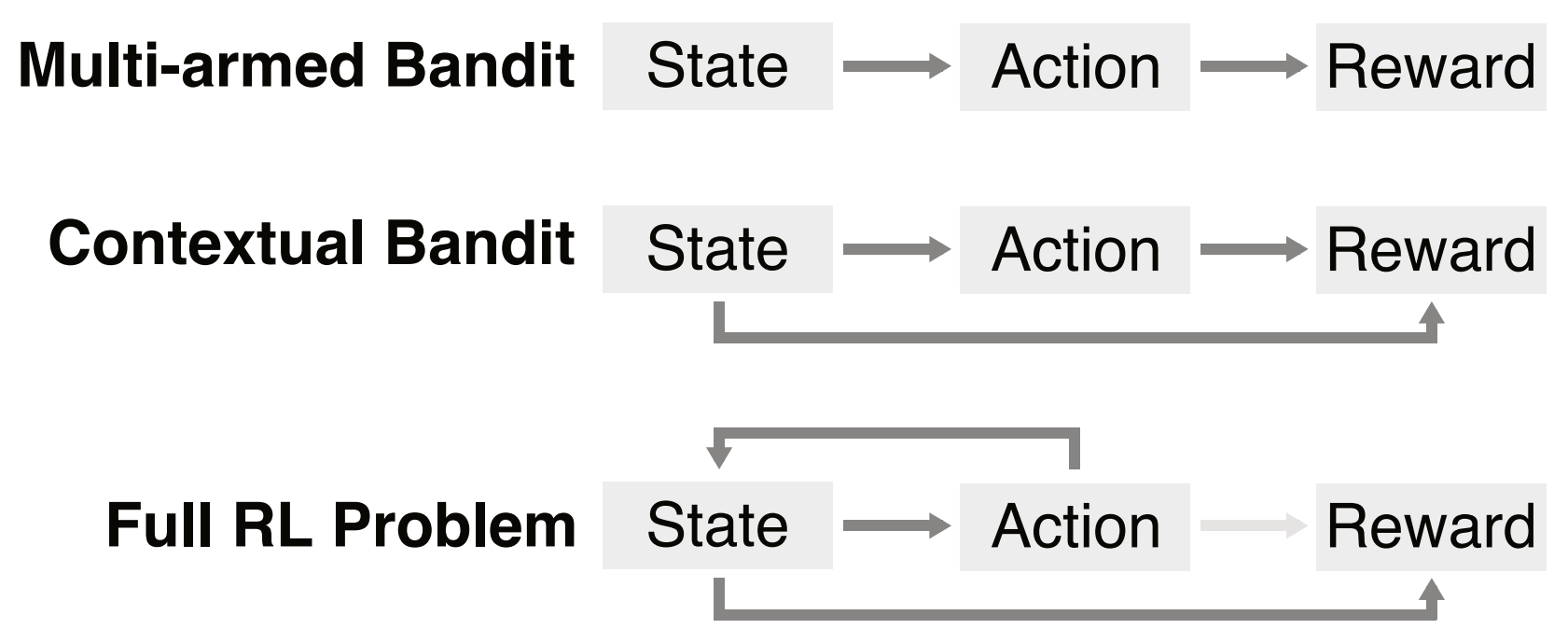
To this end we prepared a public dataset of drug vulnerability measurements from >1000 cancer cell lines for benchmarking of contextual bandit agents. The dataset contains genomic information for every cell line and complete drug response observations.



1,001 human cancer cell lines
11,289 patients
Filter by clinical prevalence
Mutations
Amplifications and deletions
Promoter Hypermethylations
Clinically relevant genomic features for sensitivity modeling
Screen against 265 compounds
cell survival %
drug concentration

Iorio et al., 2016, A Landscape of Pharmacogenomic Interactions in Cancer

The bandit problem's roots are in the rows of “one-armed bandit” slot machines seen in casinos. Each machine has a different probability of a payout and your goal is to maximize the total payout. You are limited by both the total number of bandits you can pull in a fixed period of time and uncertainty regarding which machine will deliver the best payout. The bandit problem here involves a tradeoff between exploration and exploitation

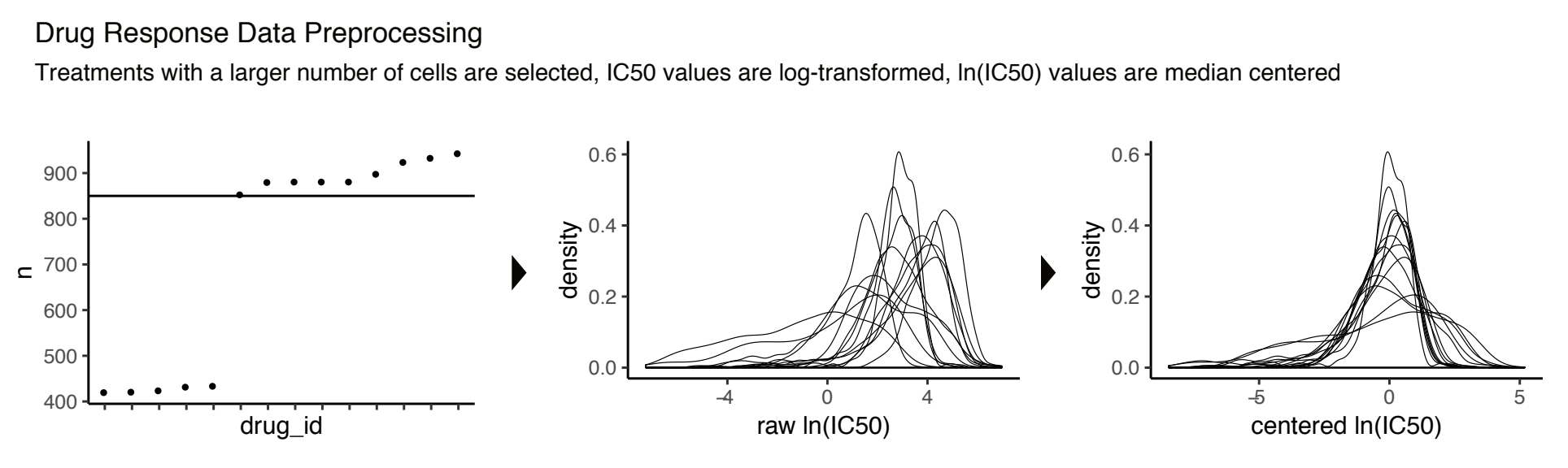

Multi-armed Bandit State → Action → Reward

Contextual Bandit State → Action → Reward

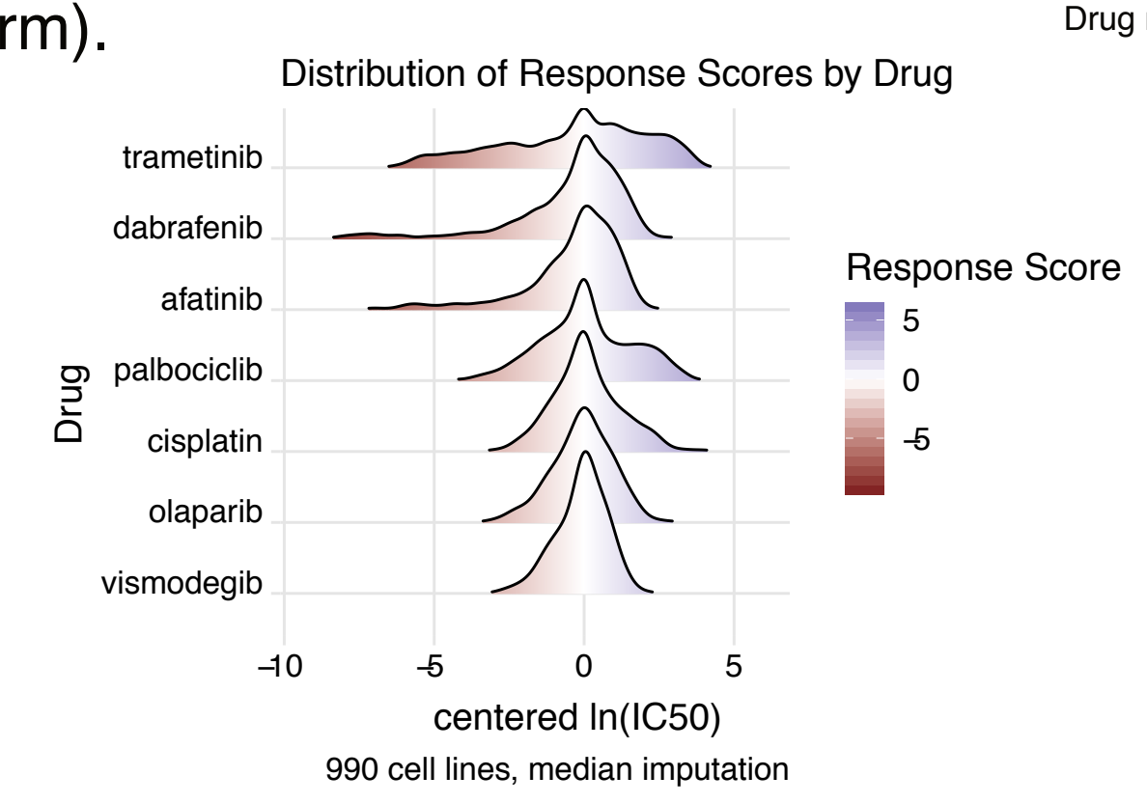
Full RL Problem State → Action → Reward

Methods - Dataset synthesis

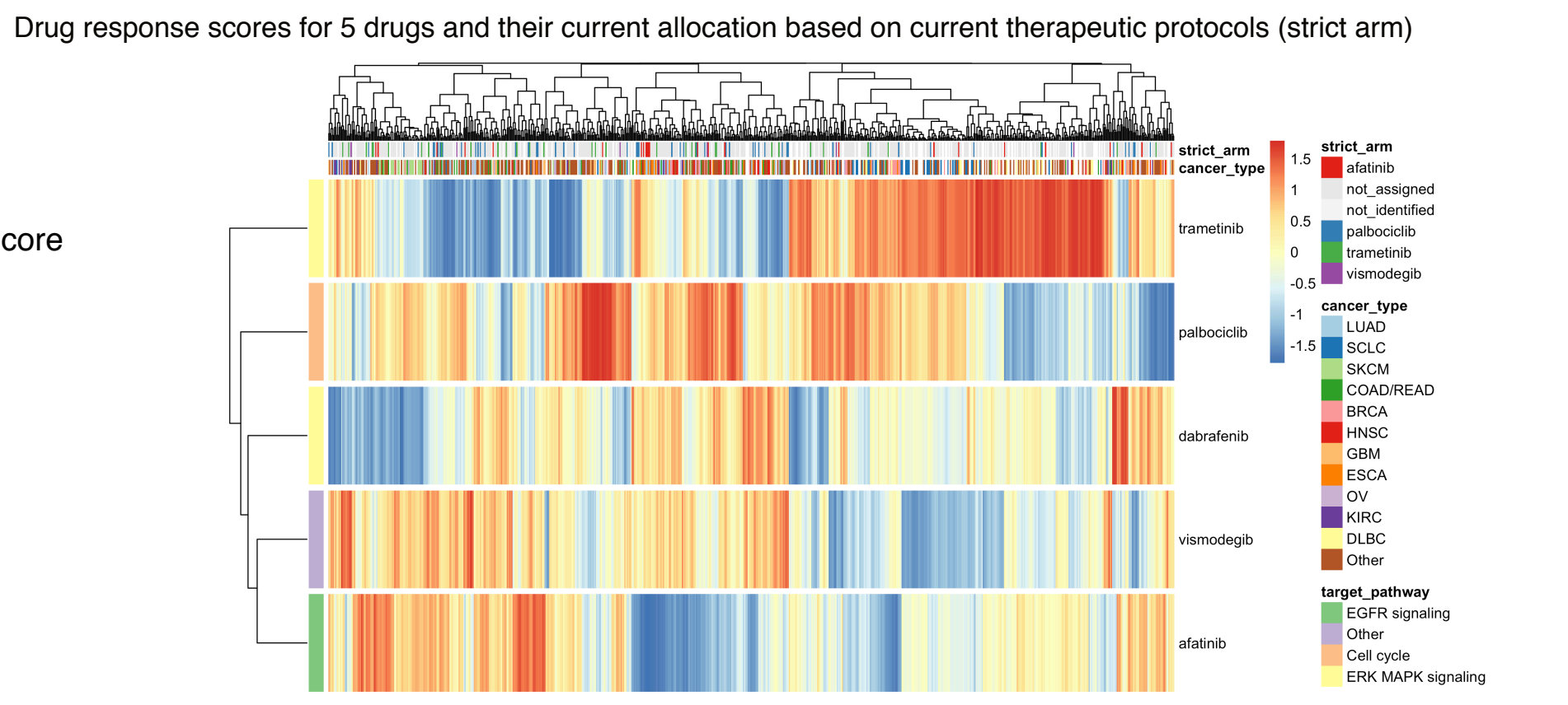
Next to data pre-processing, we searched clinical guidelines and trial protocols to identify FDA approved targeted therapeutics that have genetic biomarkers of an evidence level >2A. We defined a set of therapeutic protocols that follow a "IF gene X is altered THEN administer drug Y" structure (strict arm).


Drug Response Data Preprocessing

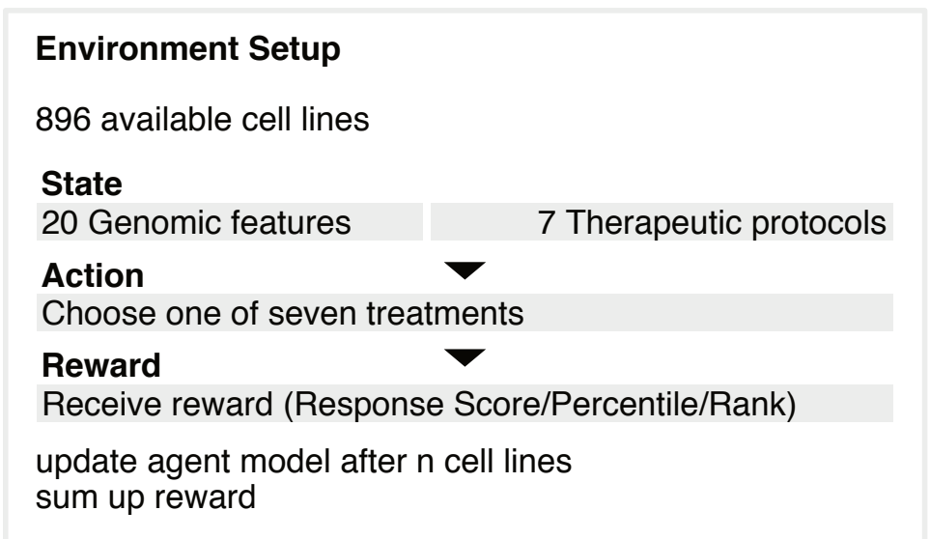
Treatments with a larger number of cells are selected, IC50 values are log-transformed, ln(IC50) values are median centered


Distribution of Response Scores by Drug

990 cell lines, median imputation


Drug response scores for 5 drugs and their current allocation based on current therapeutic protocols (strict arm)

Methods - Experimental Design


Environment Setup

896 available cell lines

State
20 Genomic features | 7 Therapeutic protocols

Action
Choose one of seven treatments

Reward
Receive reward (Response Score/Percentile/Rank)

update agent model after n cell lines
sum up reward

896 cancer cell lines are matched to one of seven available treatments subsequently. Each action is followed by a reward.

Therapeutic Protocols:

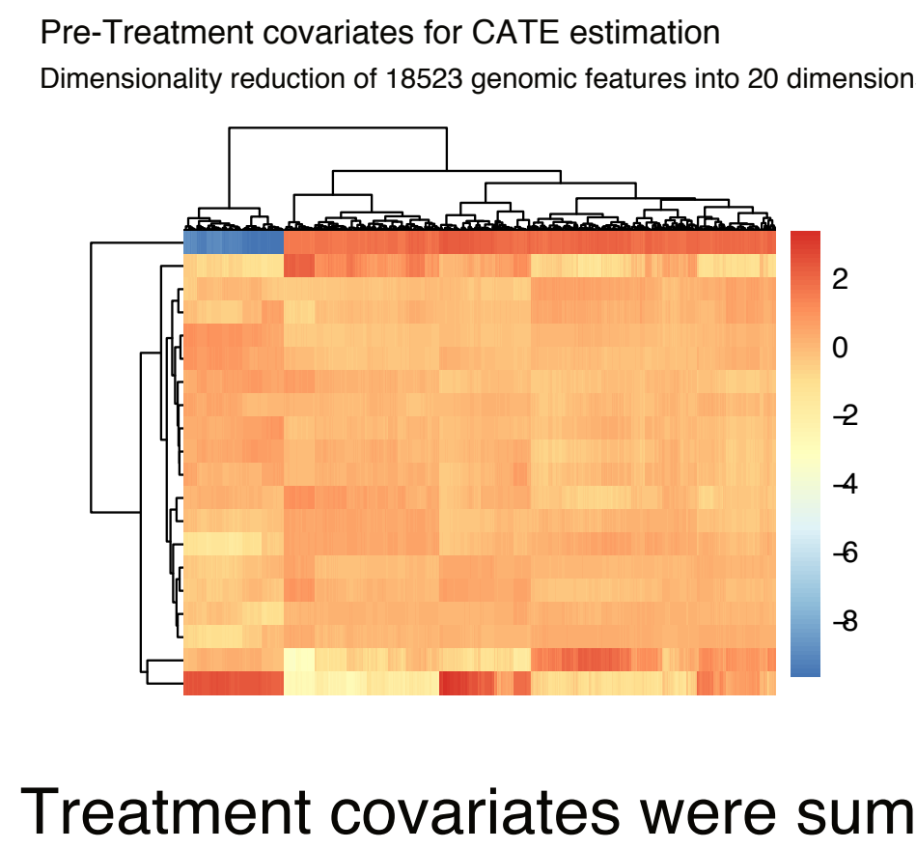
Trametinib - GNA11, NF1, BRAF non-V600E
Dabrafenib - BRAF V600E
Vismodegib - PTCH1
Afatinib - EGFR, ERBB2
Palbociclib - Rb expression & CCND1/ CDK4 amplification
Olaparib - BRCA1, BRCA2

Therapeutic Protocols applied to genomic data

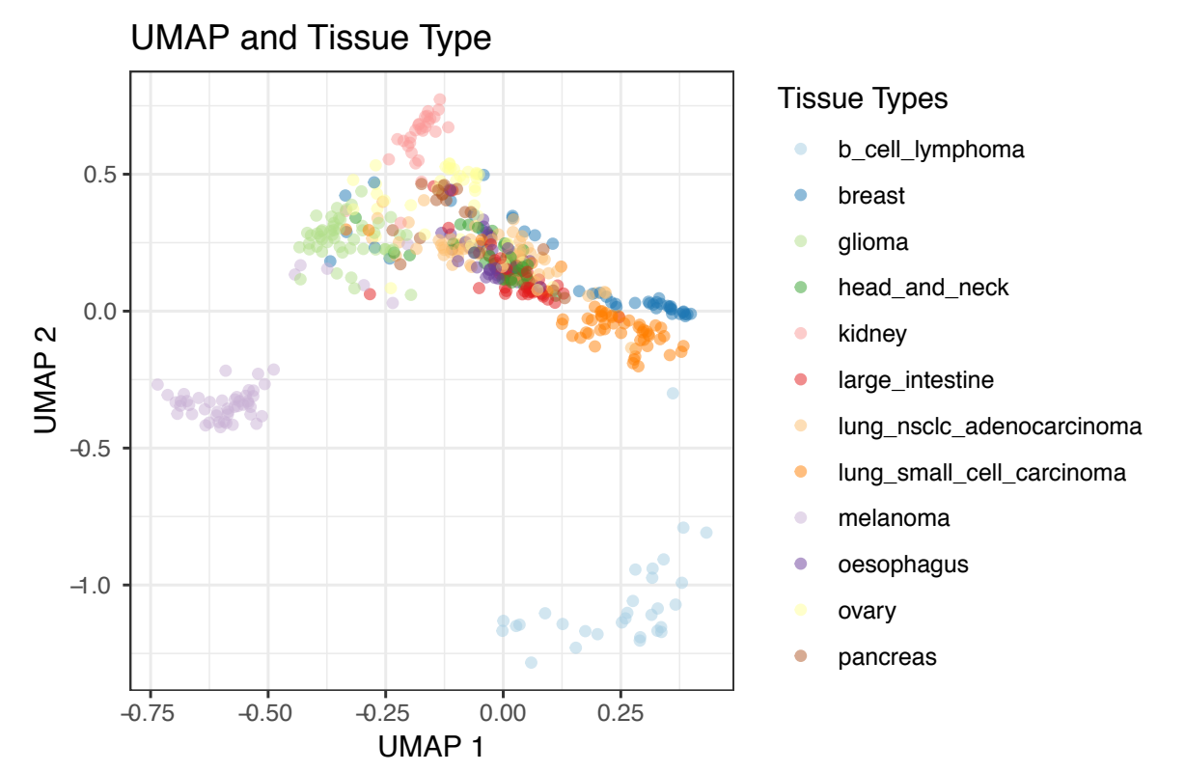
To guide treatment decision making, the agent is provided with 20 genomic features (right) and prior knowledge (bottom left), formalized in one-hot encoded current therapeutic protocols.

Based on this information the agent chooses one of seven treatments and received a reward proportional to the drug response.

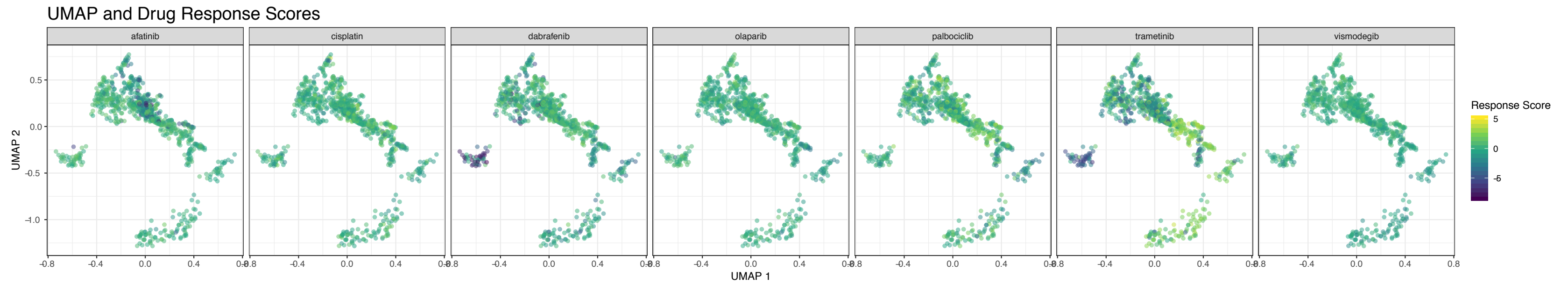
The agent's model is updated iteratively. We compared multiple models and reward functions.


Pre-Treatment covariates for CATE estimation

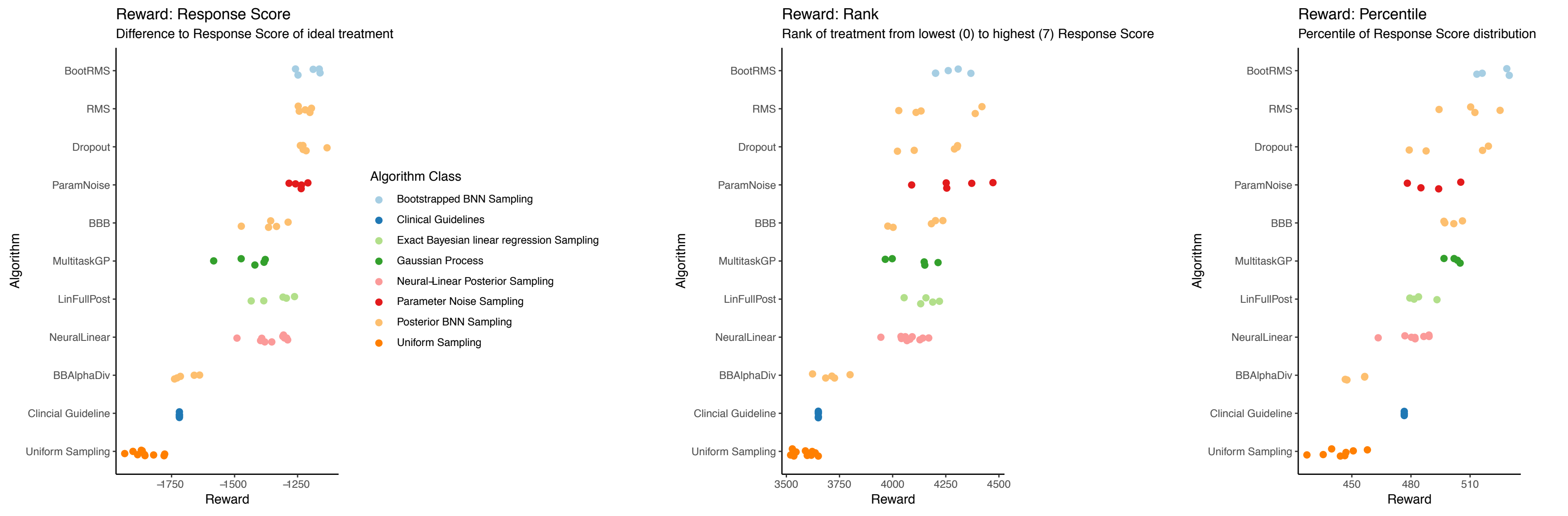
Dimensionality reduction of 18523 genomic features into 20 dimensions


UMAP and Tissue Type

Treatment covariates were summarized using uniform manifold approximation and projection (UMAP). The Treatment covariates included tissue type, mutation status, CNVs and gene expression. UMAP recovered tissue type while not directly recovering overall drug sensitivities.


UMAP and Drug Response Scores

Results - Contextual Bandit Agents outperform current therapeutic protocols in-silico


Reward: Response Score
Difference to Response Score of ideal treatment

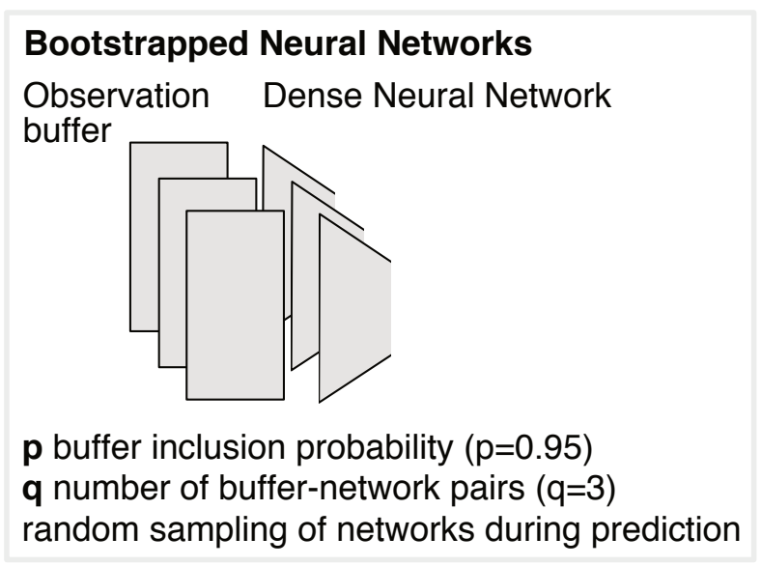
Reward: Rank
Rank of treatment from lowest (0) to highest (7) Response Score

Reward: Percentile
Percentile of Response Score distribution

We further evaluated alternative reward functions, such as the rank of each treatment option (from 1 to 7) or the percentile of the Response Score (from 0 to 1).

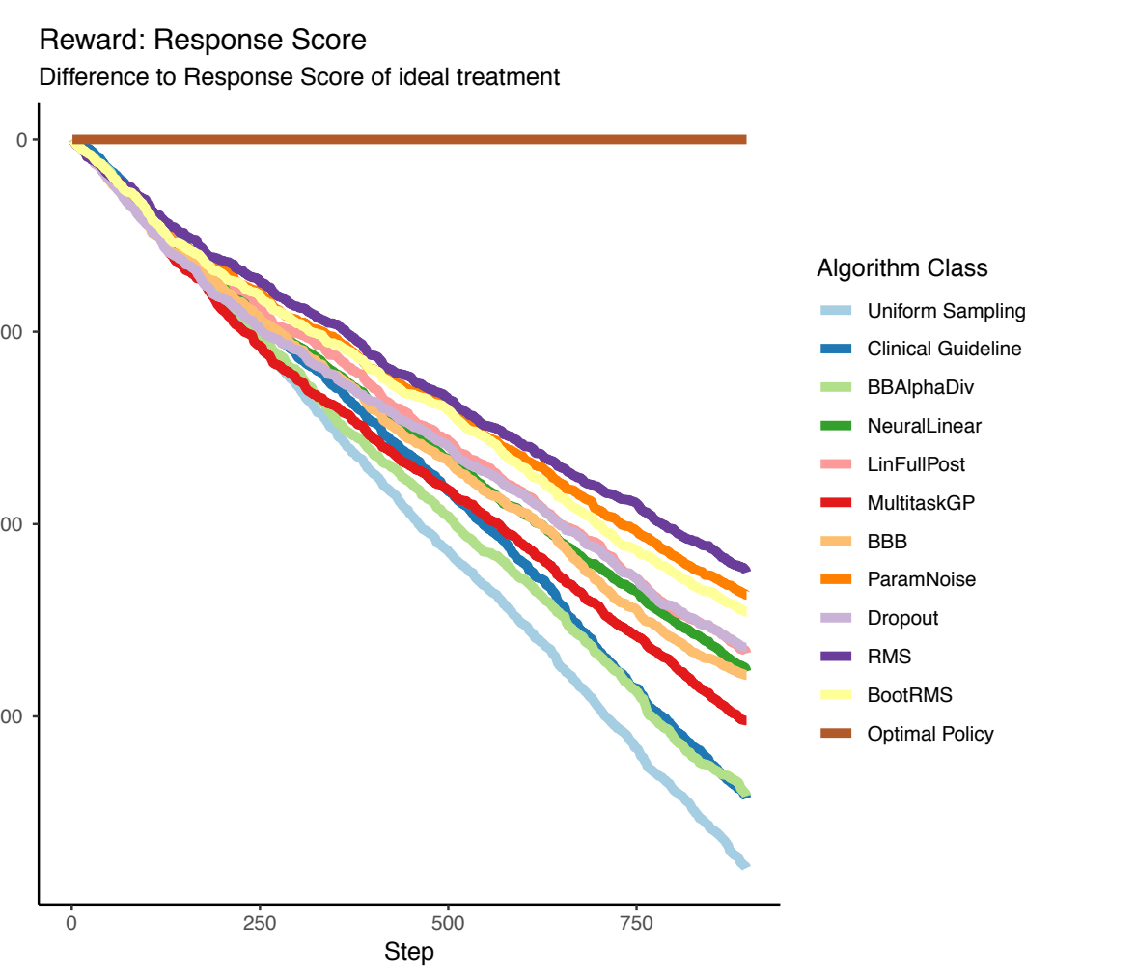
We evaluated different classes of models in the our contextual bandit environment. Strict adherence to current clinical guidelines, as codified in the therapeutic protocols, consistently outperformed random allocation of treatments. However, most models were able to outperform agents that adhered to therapeutic protocols only.

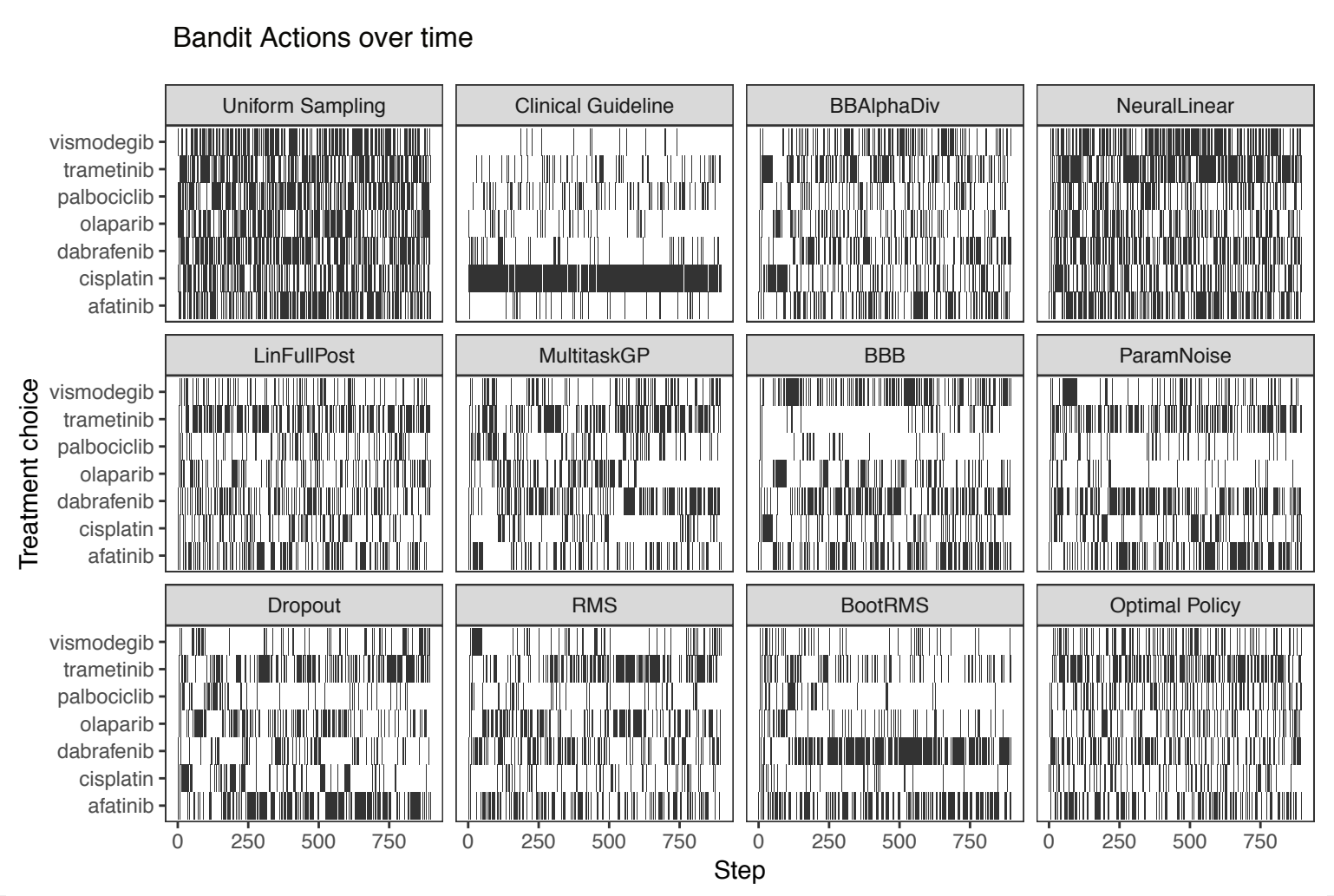
Three Neural Network algorithms, bootstrapped-, greedy and Dropout based networks, consistently scored higher rewards compared to linear methods or Gaussian Processes, independent of the reward function.


Bootstrapped Neural Networks

Observation buffer | Dense Neural Network

p buffer inclusion probability (p=0.95)
q number of buffer-network pairs (q=3)
random sampling of networks during prediction

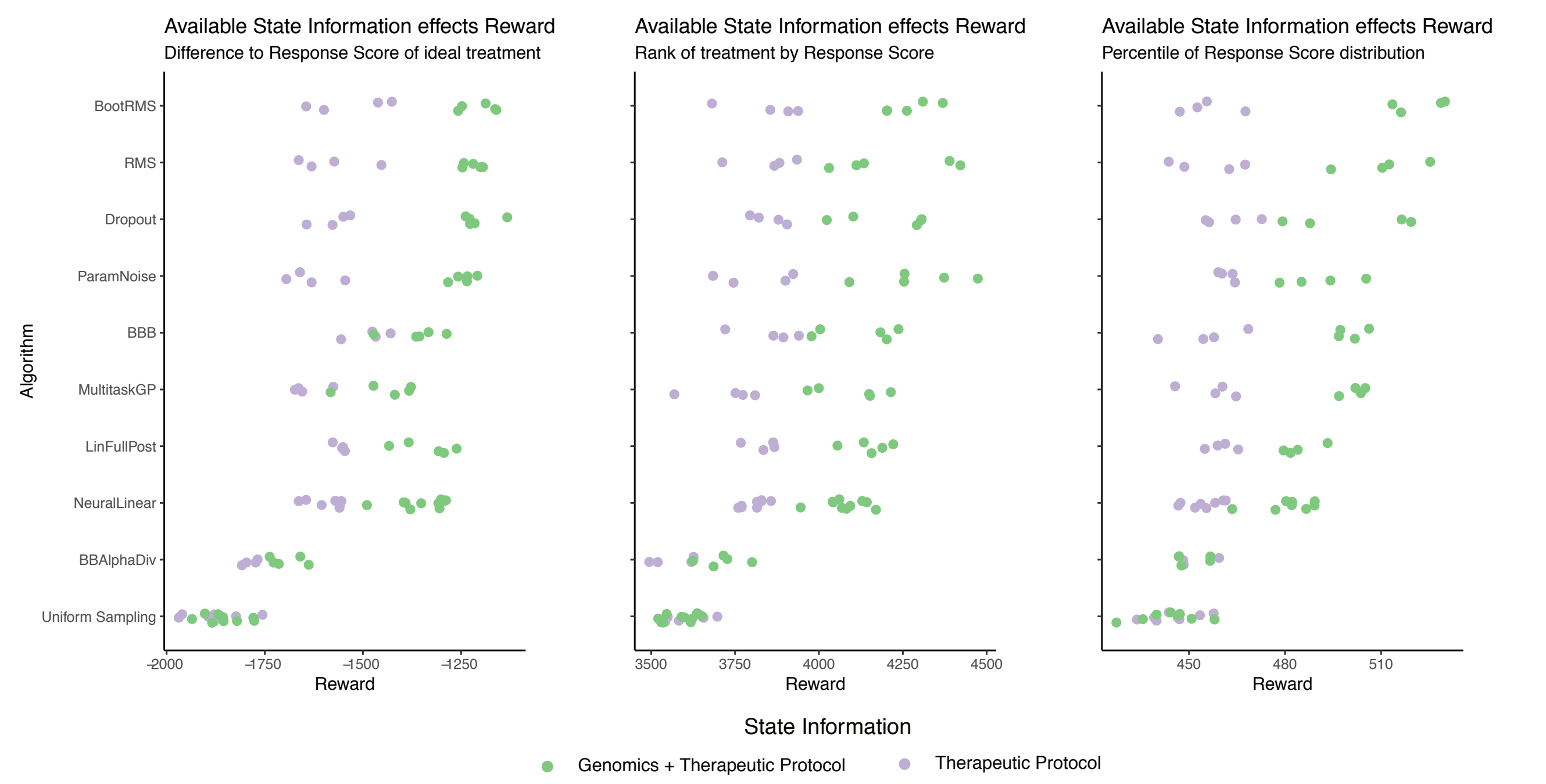

Reward: Response Score
Difference to Response Score of ideal treatment


Bandit Actions over time

Results - Agent Performance depends on genomic features

Next, we evaluated the agent performance in a scenario with only therapeutic protocol assignments as state information. Most agents performed systematically better in environments with available genomic information, independent of reward function.

Additional experiments with only genomic data as state information confirmed this observation (data not shown).


Available State Information effects Reward
Difference to Response Score of ideal treatment

Available State Information effects Reward
Rank of treatment by Response Score

Available State Information effects Reward
Percentile of Response Score distribution

Conclusion

Assignment mechanisms in precision oncology programs can be framed as a contextual bandit problem.

When provided with genomic information and expert knowledge, contextual bandit agents outperform current clinical standards in a in-vitro cancer drug response dataset, in scenarios with three different reward functions. The availability of genomic information increases the performance of most agents.

Among the most successful agents were bootstrapped or simple dense neural networks that acted greedily. In principle, both bootstrapped and dropout networks add uncertainty information by sampling from multiple related models.

This study has several limitations including: (I) In-vitro drug response data of cancer models has limited transferability into a clinical context, (II) The response scores are on average lower in treatments vs. controls, (III) Cisplatin is a limited reference treatment for all considered cancer types.

In the future, we plan to validate our findings in alternative in-vitro drug response datasets, PDX experiments and pre-clinical Organoid model data. In addition, we plan to subsample the available genomic information and its impact on model performance.

We would like to stimulate an open discussion about the limitations and potential benefits of AI guided treatment assignments in precision oncology program to minimize collective treatment regret. We acknowledge that further analysis needs to focus on avoidable regret on a per-patient level.

References

Iorio et al., 2016, A Landscape of Pharmacogenomic Interactions in Cancer
Riquelme, Tucker and Snoek, 2018, Deep Bayesian Bandits Showdown
Osband et al., 2016, Deep Exploration via Bootstrapped DQN

https://github.com/NiklasTR/oncoassign

@Niklas_TR, @Mingyu07550306, @nxpatel, @kunhsingyu, @alexdamour
niklas_rindtorff@hms.harvard.edu