Machine learning-derived gene signature predicts strong sensitivity of several solid tumors to the alkylating agent LP-184

U Kathad[1], A Kulkarni[1], JP Richard[2], T Lehman[2], R Modali[2], K Bhatia[1], P Sharma[1], F Elloumi[3], Y Pommier[3], W Reinhold[3]

[1]Lantern Pharma Inc. , [2]Reprocell USA, [3]National Cancer Institute

Permanent Abstract Number: 2090

Background

- A novel clinical agent, LP-184, is being developed by Lantern Pharma in conjunction with a dedicated machine learning-guided response signature, to allow optimal benefit and positioning of LP-184 through genomics-guided therapy.
- We report drug response predictions using RADR® (Response Algorithm for Drug Positioning and Rescue), a proprietary artificial intelligence (AI)-driven platform, and CellMinerCDB (cross database)™ [1], a systems biology platform integrating molecular and pharmacological datasets on cancer cell lines.

Challenges

- There are many approved and non-approved drugs that are focused toward one or few indications, which limits their use as potential multi-cancer or combination agents. On the other hand, objectives of precision oncology include identifying patient responsiveness to a given treatment and prevent potential mistreatments through molecular profiling. In both the cases, predictive gene expression-based candidate biomarkers are a promising and practical means to this purpose.
- It has been difficult to derive gene signatures that predict drug response accurately using completely unseen data specially with a small number of training records. Most often, the machine learning based models perform well on the validation set but not in the completely blind set.

Data types

- For model building and prediction testing, we used NCI60 cell line drug sensitivity (IC50) data and gene expression data.
- We further analyzed the gene signature-based correlation of the multi-omic features such as mutation, methylation, copy number variation (CNV) and protein expression.
- The laboratory experimental studies on completely new set of cell lines were used to test the performance of the signatures.
- Cancer Cell Line Encyclopedia (CCLE) [2] gene expression data on a total 1036 cell lines covering 22 cancer types and subtypes were used to extrapolate and predict LP-184 sensitivity in order to find the potential new target indication.

Objectives

1) Find potential expression-based gene signature for LP-184 that can predict drug response using independent machine learning driven platforms: CellMinerCDB and RADR®, including derivation of pan-cancer and solid tumor-specific gene signatures, and compare their performance with real world data

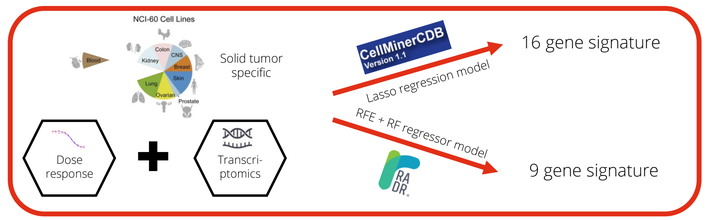
2) Identify potential new target indications for LP-184 using CCLE gene expression database

3) Annotate the relevance of signature genes showing multi-omics correlations with LP-184 sensitivity

Pan-cancer signature development

We used drug sensitivity and gene expression data from 58 NCI-60 panel cell lines. A CellMinerCDB lasso regression model generated a 38 gene signature from the input of ~23000 genes.

Solid tumor-specific signature development



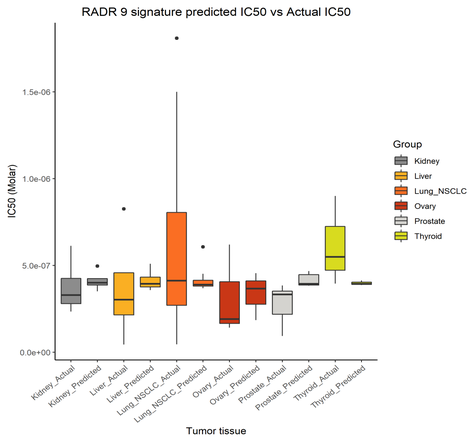
- We used CellMinerCDB and RADR® to build two different solid tumor specific models. In the development of both the model, we considered 52 solid tumor cell lines. We used RADR platform to derive the smallest gene set, that can predict with high accuracy. RADR used multiple sets of filtration process (statistical and biological filtering) to narrow down ~23000 genes (full transcriptome) to ~100 genes for modeling input.
- After identifying filtered top ranked 100 genes, RADR® used RFE (Recursive Feature Elimination) to reduce the gene set even further down. A Random Forest (RF) modeling with 10-fold CV (Cross Validation) and random search tuning method provided a final 9 gene signature.

Gene expression-based machine learning signatures can be used to predict cancer drug sensitivity, identify new indications and provide MoA insights.

Comparison of signature performance

#	Signature details	Number of genes	Cell Line IC50 Prediction Accuracy (%) using 4-fold cutoff	Cell Line IC50 Prediction Accuracy (%) using 2-fold cutoff
1	Cell Miner – Lasso – Regression – all cancer types	38	29/37 (78.37%)	17/37 (45.94%)
2	Cell Miner – Lasso – Regression – solid tumors	16	31/37 (83.78%)	18/37 (48.64%)
3	RADR® – RFE Random Forest – Regression – solid tumors	9	33/37 (89.18%)	20/37 (54.05%)

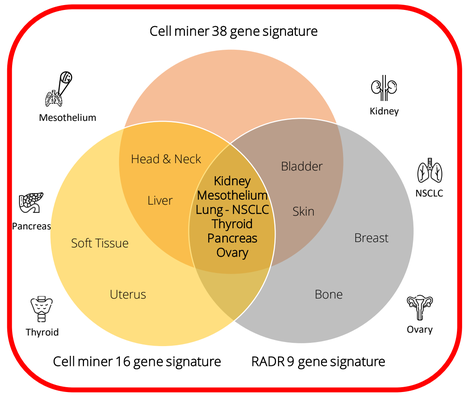
Total number of unique integrated signature genes: 56



- We applied these above three developed and tuned model to the laboratory experimental study, which is completely new set / blind set to test the performance.
- From the wet lab study, we generated drug sensitivity on a total of 37 cell lines, in which we tested our trained and validated model.
- From the above table, we were able to achieve more than 80% of accuracy for the solid tumor specific models, considering accurate prediction if the predicted IC50 is < (+/-) 4 fold change compared to actual IC50. With the stronger cut-off of < (+/-) 2 fold change, the best model predicted with more than 50% accuracy.
- The RADR derived 9 gene signature gave closest IC50 prediction to actual IC50 data, compared to other signatures.

Predicted cancer indications for LP-184

Top 10 indication having smallest median predicted IC50



We predicted drug sensitivity of entire CCLE cell line set (1036) using all the three signature sets. Based on the predicted tissue median IC50, we made a list of top 10 indications having smallest IC50 and top 10 indications having largest IC50 (after removing indications having < 10 cell lines). We identified Kidney, Mesothelium, Lung – NSCLC, Thyroid, Pancreas and Ovary as top common sensitive indications among all three signature predictions.

On the other hand, Lymph, Blood, Lung – SCLC, Bowel, Brain – CNS and Esophagus were top common insensitive indications.

Multi-omics correlates of LP-184 sensitivity

Expression patterns of almost all the machine learning-derived signature genes are significantly correlated (p value < 0.05) with LP-184 sensitivity. We present below, a snapshot of selected signature genes having significant multi-omic correlations with LP-184 sensitivity.

#	Signature gene	Expression (R^2)	Methylation (R^2)	CNV (R^2)	Hypothesis of gene involvement in the MoA of LP-184
1	APP	0.739	-0.531		Targeting of APP can induce apoptosis
2	SMARCC1	-0.520		-0.340	Decreased expression hampers DNA Damage Response (DDR) that would otherwise repair LP-184-DNA adducts via TC-NER
3	EGFR	0.638	-0.554		Inhibition by acylfulvene moiety leads to impaired proliferation signaling
4	NEK6	0.697	-0.505		Targeting of Nek6-dependent DNA damage checkpoint activity is required for proper cell cycle arrest upon DNA damage
5	SLC43A3	-0.403	0.412	-0.257	Decreased transporter levels may reduce drug efflux
6	SQSTM1	0.611	-0.488		Connected with other signature genes APP, EGFR and AHR, targeting of SQSTM1 can promote cell cycle arrest and apoptosis
7	AHR	0.524			Targeting of AHR can promote cell cycle arrest and apoptosis
8	FANCE	-0.362	-0.391		Decreased expression hampers DDR that would otherwise repair LP-184-DNA adducts via TC-NER
9	POLR1C	-0.410		-0.270	Decreased expression hampers DDR that would otherwise repair LP-184-DNA adducts via TC-NER
10	PTPRC	-0.681		-0.308	Decreased expression may augment LP-184 sensitivity in the background of a less aggressive cancer phenotype
11	HIST1H2AM	-0.688	0.620		Decreased expression impairs chromatin modeling that in turn regulates other gene expression patterns
12	RHOH	-0.704		-0.272	Decreased expression may augment LP-184 sensitivity in the background of a less aggressive cancer phenotype
13	BCAR3	0.630	0.563		Targeting of BCAR3 can promote cell cycle arrest and apoptosis

R^2 = Pearson Correlation coefficient, CNV = Copy Number Variation, MoA = Mechanism of Action

#	R^2 - correlation coefficient range (+/-)	Interpretation (positive / negative)
1	0 to 0.33	No to weak correlation
2	0.33 – 0.66	Moderate to strong correlation
3	0.66 – 1	Strong to perfect correlation

From the annotations, we found many signature genes (POLG2, SMARCC1, RDM1, MEN1, NEK6, WRN, WDR48, CHEK1, SMARCB1, MRE11A, UVRAG, BCAS2, FANCE, SMARCB1, POLR1C, SMARCC1, SLFN11, MRE11A, NKX3-1) related to DNA Damage Response and few genes (POLD1, HIST1H2AM, AFF4, PTPN7) related to chromatin modeling, which overlap with the proposed MoA of LP-184 [1, 3].

In conclusion, our results demonstrate that LP-184 development guided by tumor gene expression patterns modeled using a combination of algorithms and signatures provides a valuable component to the armamentarium of drugs in diverse solid tumors.

References

1) https://discover.nci.nih.gov/cellminercdb/
2) https://portals.broadinstitute.org/ccle
3) https://www.genecards.org/

Contact

Umesh Kathad, MS
Bioinformatics & Platform Development Manager
umesh@lanternpharma.com