# Drug Discovery

## Application of artificial intelligence to predict a new class of novel synthetic lethal targets.

Abstract No: 2598

**Background:** Synthetic lethal targets are proteins that are contextually vulnerable. Inhibitors of PARP1, for example, selectively produce a lethal phenotype in the context of cancer cells which have lost BRCA1 or BRCA2 function. As a high mutation rate is a hallmark of many cancers, targeting synthetic lethal interactions to selectively inhibit cancer cells with altered genetic backgrounds may increase the specificity and efficacy of therapeutics. Recently, clinical trials have targeted synthetic lethal pairs such as EGFR and BRAF, TP53 and BCL2, and PTEN and CHD1. Previous attempts to identify synthetic lethal targets have relied on empirical results from published studies of biological pathways perturbed in cancer cells. Developing strategies to rapidly identify synthetic lethals by combining multiple experimental and computational approaches would result in a new class of potential cancer drug targets beyond the existing efforts that rely on single experimental or computational methods alone.

**Methods:** Here we present Expansive AI, an artificial intelligence augmented knowledge network that enables rapid hypothesis generation for accelerated discovery research. Using a purpose-built, hypergraph database of massive, integrated genomic and biomedical data, we can query all synthetic lethals and their component genes, as well as a wealth of data related to these genes. The database of biological data includes 11,000+ cancer genomes from TCGA, prior knowledge resources such as gene ontology and pathway resources, and experimental data including chemical and protein interaction and patent data. The hypergraph's architecture allows for linking and nesting data, enabling efficient extraction of biologically relevant features.

**Results:** Using these features, a neural network classified 540 new candidate pairs that have previously not been reported. The candidate pairs were filtered to include only known oncogenes and least-studied genes. This produced a list of gene pairs which may represent the most novel class of synthetic lethal target candidates identified to date.

**Conclusions:** We highlight the results of this AI-based approach and discuss validation efforts of the predicted interactions in specific cancer contexts.

# Clinical Diagnosis & Prognosis

## Machine learning algorithm analysis using a commercial 592-gene NGS panel to accurately predict tumor lineage for carcinoma of unknown primary (CUP).

Abstract No: 3083

**Background:** The diagnosis of a malignancy is typically informed by clinical presentation and tumor tissue features including cell morphology, immunohistochemistry, cytogenetics, and molecular markers. However, in approximately 5-10% of cancers, ambiguity is high enough that no tissue of origin can be determined, and the specimen is labeled as a Cancer of Occult\Unknown Primary (CUP). Lack of reliable classification of a tumor poses a significant treatment dilemma for the oncologist leading to inappropriate and/or delayed treatment.

**Methods:** 40,000 tumor patients with NGS data were used to construct a multiple parameter lineage-specific classification system using an advanced machine learning approach. The dataset for each classifier was split 50% for training and the other 50% for testing. The training task for each classifier was to identify the cases that were similar to the cases it was trained on against a backdrop of randomly selected cases of other histological origins.

**Results:** Tumor lineage classifiers predicted the correct classifications where the primary site was known with accuracies ranging between 85% and 95%. When applied to CUP cases (n = 500), an unequivocal result could be obtained 100% of the time.

**Conclusions:** Lineage predictors can render a histologic diagnosis to CUP cases that can inform treatment and potentially improve outcomes.

## Evaluating the capacity of connectome analysis to predict survival in high-grade astrocytoma.

Abstract No: 2049

**Background:** While factors such as age, histology and tumor molecular variants (e.g. IDH status) contribute to prognosis in patients with high grade astrocytoma (HGA), there remains a wide variability in patient survival outcomes. The connectome, or brain network organization, incorporates biologic, molecular and environmental processes providing a uniquely parsimonious summary of key prognostic factors. This study compared the capacity of machine learning (ML) models based on baseline connectomics and clinical variables to predict patient survival in HGA.

**Methods:** Patients with a new diagnosis of HGA and a presurgical 3D, T1-weighted MRI available were retrospectively identified. Individual patient connectomes were derived from MRI with 90 cortical/subcortical features. Presurgical clinical features included age, gender, histology, tumor grade and IDH status. Three ML algorithms were implemented: extreme learning machine with Buckley–James estimator (ELMBJ), random survival forest (RSF) with logrank splitting and RSF with concordance index (CI) splitting. For each algorithm, we used a 60/40 training/testing split with 50 iterations and CI as the performance metric. We tested three models: 1) connectome only, 2) clinical only, and 3) connectome plus clinical variables.

**Results:** Of patients identified (n = 105), 66 had glioblastoma and 39 had anaplastic astrocytoma. Thirty-eight harbored IDH mutation. Median overall survival was 27.43 months (SD 39.57). Connectome-only models showed better prediction performance compared to clinical-only models across all algorithms. ELMBJ showed the best performance (connectome median CI = 0.522, clinical CI = 0.201). Connectome models also performed as well as combined models (e.g. median CI = 0.523 for ELMBJ).

**Conclusions:** This study demonstrates the potential of a connectome model to predict survival of patients with HGA. Replication in a larger sample is required to validate these results and refine ML models including examination of additional clinical features. If successful, use of a simple T1 MRI could provide additional variables to augment existing prognostic prediction, especially in scenarios where tumor genotyping is not available.

## A predictive model for survival in non-small cell lung cancer (NSCLC) based on electronic health record (EHR) and tumor sequencing data at the Department of Veterans Affairs (VA).

Abstract No: 109

**Background:** Machine learning tools based on EHR data hold promise to help avoid unnecessary risks associated with lung cancer and its treatment. Additionally, molecular genetic profiling is becoming an integral tool for clinicians to individualize treatment for lung cancer. However, relatively few survival models have been built that integrate this data in individualized predictive models. Here, we combine real-world EHR and tumor sequencing data from the VA Precision Oncology Data Repository (PODR) to build accurate individualized survival predictions in newly-diagnosed NSCLC patients.

**Methods:** We identified a cohort of 356 VA patients newly diagnosed with NSCLC for whom EHR, cancer registry, and targeted tumor sequencing data is available in PODR. We defined 41 features reflecting 15 baseline clinical and demographic characteristics from the EHR and registry, such as age, race, stage, histology, and therapy. We also defined features reflecting 206 clinically actionable somatic variants. We selected 5 important variants for inclusion in the model, as well as the total number of mutations. We trained a random forests algorithm to predict 1-year survival. Precision, recall, and area under the ROC curve (AUC) were assessed using 5-fold cross validation.

**Results:** Mean age at diagnosis was 66 years. The majority of patients had late stage disease (15% stage I, 6% II, 15% III, 44% IV), and 59% of patients received systemic therapy. 45% died within 1 year of diagnosis, and 55% survived past 1 year. Our predictive model for 1-year survival achieves strong results. Cross-validated AUC is 0.79 (SD 0.08), precision is 0.79 (SD 0.07), recall is 0.74 (SD 0.07), suggesting that the trained model combining clinical and genomic features is effective at predicting 1-year survival.

**Conclusions:** By integrating real-world EHR and sequencing data, we built a highly accurate predictive model of 1-year survival in NSCLC patients at the VA. Such a model, after ongoing validation in a larger cohort, offers the ability to make individualized predictions that could inform patient care to improve outcomes

# Treatment Response

## A machine learning based prediction model of anti-PD-1 therapy response using noninvasive clinical information and blood markers of lung cancer patients.

Abstract No: e14138

**Background:** Immune checkpoint inhibitors have become breakthrough therapy for various types of cancers. However, regarding their total response rate around 20% based on clinical trials, predicting accurate aPD-1 response for individual patient is unestablished. The presence of PD-L1 expression or tumor infiltrating lymphocyte may be used as indicators of response but are limited. We developed models using machine learning methods to predict the aPD-1 response.

**Methods:** A total of 126 advanced NSCLC patients treated with the aPD-1 were enrolled. Their clinical characteristics, treatment outcomes, and adverse events were collected. Total clinical data (n = 126) consist of 15 variables were divided into two subsets, discovery set (n = 63) and test set (n = 63). Thirteen supervised learning algorithms including support vector machine and regularized regression (lasso, ridge, elastic net) were applied on discovery set for model development and on test set for validation. Each model were evaluated according to the ROC curve and cross-validation method. Same methods were used to the subset which had additional flow cytometry data (n = 40).

**Results:** The median age was 64 and 69.8% were male. Adenocarcinoma was predominant (69.8%) and twenty patients (15.1%) were driver mutation positive. Clinical data set (n = 126) demonstrated that the Ridge regression (AUC: 0.79) was the best model for prediction. Of 15 clinical variables, tumor burden, age, ECOG PS and PD-L1, were most important based on the random forest algorithm. When we merged the clinical and flow cytometry data, the Ridge regression model (AUC:0.82) showed better performance compared to using clinical data only. Among 52 variables of merged set, the top most important immune markers were as follows: $CD3^+CD8^+CD25^+/T_{eff}\text{-}CD28$, $CD3^+CD8^+CD25^-/T_{eff}\text{-}Ki\text{-}67$, and $CD3^+CD8^+CD25^+/T_{eff}\text{-}NY\text{-}ESO/T_{eff}\text{-}PD\text{-}1$, which indicate activated tumor specific T cell subset.

**Conclusions:** Our machine learning based model has benefit for predicting aPD-1 responses. After further validation in independent patient cohort, the supervised learning based non-invasive predictive score can be established to predict aPD-1 response.

## Using machine learning algorithms to predict response and toxicity to ICIs in melanoma patients.

Abstract No: 2581

**Background:** There is growing interest in optimizing patient selection for treatment with immune checkpoint inhibitors (ICIs). We postulate that phenotypic features present in metastatic melanoma tissue reflect the biology of tumor cells, immune cells, and stromal tissue, and hence can provide predictive information about tumor behavior. Here, we test the hypothesis that machine learning algorithms can be trained to predict the likelihood of response and/or toxicity to ICIs.

**Methods:** We examined 124 stage III/IV melanoma patients who received anti-CTLA-4 (n = 81), anti-PD-1 (n = 25), or combination (n = 18) therapy as first line. The tissue analyzed was resected before treatment with ICIs. In total, 340 H&E slides were digitized and annotated for three regions of interest: tumor, lymphocytes, and stroma. The slides were then partitioned into training (n = 285), validation (n = 26), and test (n = 29) sets. Slides were tiled (299x299 pixels) at 20X magnification. We trained a **deep convolutional neural network** (DCNN) to automatically segment the images into each of the three regions and then deconstruct images into their component features to detect non-obvious patterns with objectivity and reproducibility. We then trained the DCNN for two classifications: 1) complete/partial response versus **progression of disease** (POD), and 2) severe versus no immune-related adverse events (irAEs). Predictive accuracy was estimated by area under the curve (AUC) of **receiver operating characteristics** (ROC).

**Results:** The DCNN identified tumor within LN with AUC 0.987 and within ST with AUC 0.943. Prediction of POD based on ST-only always performed better than prediction based on LN-only (AUC 0.84 compared to 0.61, respectively). The DCNN had an average AUC 0.69 when analyzing only tumor regions from both LN and ST data sets and AUC 0.68 when analyzing tumor and lymphocyte regions. Severe irAEs were predicted with limited accuracy (AUC 0.53).

**Conclusions:** Our results support the potential application of machine learning on pre-treatment histologic slides to predict response to ICIs. It also revealed their limited value in predicting toxicity. We are currently investigating whether the predictive capability of the algorithm can be further improved by incorporating additional immunologic biomarkers.

## Development and validation of a deep learning model to assess tumor progression to immunotherapy.

Abstract No: e20601

**Background:** Manual application of length-based tumor response criteria is the standard-of-care for assessing metastatic tumor response. It is technically challenging, time-consuming and associated with low reproducibility. In this study, we presented a novel automatic Deep Neural Networks (DNNs) based segmentation method for assessing tumor progression to immunotherapy. Next stage, AI will assist Physicians assessing pseudo-progression.

**Methods:** A data set of 39 lung cancer patients with 156 computed tomography (CT) scans was used for model training and validation. A 3D segmentation DNN DenseSharp, was trained with an input size of on CT scans of tumor with manual delineated volume of interest (VOI) as ground truth. The trained model was subsequently used to estimate the volumes of target lesions via 16 sliding windows. We referred the progression-free survival (PFS) only considering tumor size as PFS-T. PFS-Ts assessed by longest tumor diameter (PFS-T$_{diam}$), by tumor volume (PFS-T$_{vol}$), and by predicted tumor volume (PFS-T$_{pred-vol}$) were compared with standard PFS (as assessed by one junior and one senior clinician). Tumor progression was defined as > 20% increase in the longest tumor diameter or > 50% increase in tumor volume. Effective treatment was defined as a PFS of > 60 days after immunotherapy.

**Results:** In a 4-fold cross-validation test, the DenseSharp segmentation neural network achieved a mean per-class intersection over union (mIoU) of 80.1%. The effectiveness rates of immunotherapy assessed using PFS-T$_{diam}$ (32 / 39, 82.1%), PFS-T$_{vol}$ (33/39, 84.6%) and PFS-T $_{pred-vol}$ (32/39, 82.1%) were the same as standard PFS. The agreement between PFS-T$_{vol}$, and PFS-T$_{pred-vol}$ was 97.4% (38/39). Evaluation time with deep learning model implemented with PyTorch 0.4.1 on GTX 1080 GPU was hundred-fold faster than manual evaluation (1.42s vs. 5-10 min per patient).

**Conclusions:** In this study, DNN based model demonstrated fast and stable performance for tumor progression evaluation. Automatic volumetric measurement of tumor lesion enabled by deep learning provides the potential for a more efficient, objective and sensitive measurement than linear measurement by clinicians.

# Deep learning-based predictive biomarker for immune checkpoint inhibitor response in metastatic non-small cell lung cancer.

Abstract No: 9094

**Background:** In the era of immunotherapy, immune checkpoint inhibitor (ICI) has changed the treatment paradigm in metastatic non-small cell lung cancer (NSCLC). Along with clinical trials, there is an ongoing investigation to discover the predictive biomarker of ICI which so far has unsatisfactory reliability. As an effort to enhance the predictive value of ICI treatment, we applied deep learning and developed artificial intelligent (AI) score (range from 0 to 1) to analyze the specific context of immune-tumor microenvironment (TME) extracted by scanned images from H&E slides.

**Methods:** As a groundwork, deep learning-based H&E image analyzer, Lunit SCOPE, has been trained with H&E images (n = 1824) from ICI naive NSCLC samples. For the calculation of AI score, training was conducted using responder/non-responder labeled ICI treated samples from the exploratory cohort. The ICI responder was defined as the patient with a best overall response of partial or complete response and stable disease for more than 6 months. The positivity of PD-L1 immunohistochemistry (IHC) was assessed manually by pathologists.

**Results:** The exploratory cohort is composed of NSCLC patients treated with ICI (n = 189) in Samsung Medical Center, and response to ICI was observed in 72 (38.1%) patients. Median follow-up duration was 6.8 months (6.6~8.2). Samples with PD-L1 IHC positive, defined by ≥ 1%, was observed in 138 (73.0%) patients. AI score was significant higher in the responder group (median: 0.391 vs 0.205, P = 6.14e-5), and the patients with AI score above the cut-off (0.337) showed a better response to ICI (odds ratio [OR] 3.47 P = 7.34e-5) which is higher than patients with PD-L1 ≥ 1% (OR 1.92, P = 0.069). High AI score group (n = 83) showed significantly favorable PFS compared to low AI score group (n = 106, median PFS: 5.1m vs 1.9m, hazard ratio [HR] 0.51, P = 9.6e-5) and this outcome was independent with PD-L1 status (P = 6.0e-5). In subgroup analysis, PFS of PD-L1 high / AI score high group (n = 63) had longer median PFS (6.7m) compared to both PD-L1 high / AI score low group (n = 70, 4.0m, P = 0.001) and PD-L1 low/AI score low group (n = 35, 1.9m, P = 4.0e-6). Tumor infiltrating lymphocyte (TIL) density of cancer epithelium was significantly correlated with AI score (Pearson's r = 0.310, P = 1.43e-5), which suggests that AI score may partly reflect TME represented by TIL.

**Conclusions:** The AI score by machine-learned information, extracted from H&E images without additional IHC stain, could predict responsiveness and PFS of ICI treatment independent of PD-L1 IHC positivity.

# Development and clinical validation of Lantern Pharma's AI engine: Response algorithm for drug positioning and rescue (RADR).

Abstract No: 3114

**Background:** The Response Algorithm for Drug positioning and Rescue (RADR) technology is Lantern Pharma's proprietary Artificial Intelligence (AI)-based machine learning approach for biomarker identification and patient stratification. RADR is a combination of three automated modules working sequentially to generate drug- and tumor type-specific gene signatures predictive of response.

**Methods:** RADR integrates genomics, drug sensitivity and systems biology inputs with supervised machine learning strategies and generates gene expression-based responder/ non-responder profiles for specific tumor indications with high accuracy, in addition to identification of new correlations of genetic biomarkers with drug activity. Pre-treatment patient gene expression profiles along with corresponding treatment outcomes were used as algorithm inputs. Model training was typically performed using an initial set of genes derived from cancer cell line data when available, and further applied to patient data for model tuning, cross-validation and final gene signature development. Model testing and performance computation were carried out on patient records held out as blinded datasets. Response prediction accuracy and sensitivity were among the model performance metrics calculated.

**Results:** On average, RADR achieved a response prediction accuracy of 80% during clinical validation. We present retrospective analyses performed as part of RADR validation using more than 10 independent datasets of patients from selected cancer types treated with approved drugs including chemotherapy, targeted therapy and immunotherapy agents. For an instance, the application of the RADR program to a Paclitaxel trial in breast cancer patients could have potentially reduced the number of patients in the treatment arm from 92 unselected patients to 24 biomarker-selected patients to produce the same number of responders. Also, we cite published evidence correlating genes from RADR derived biomarkers with increased Paclitaxel sensitivity in breast cancer.

**Conclusions:** The value of RADR platform architecture is derived from its validation through the analysis of about ~17 million oncology-specific clinical data points, and ~1000 patient records. By implementing unique biological, statistical and machine learning workflows, Lantern Pharma's RADR technology is capable of deriving robust biomarker panels for pre-selecting true responders for recruitment into clinical trials which may improve the success rate of oncology drug approvals.

## Prediction of treatment (tx)-induced fatigue in breast cancer (BC) patients (pts) using machine learning on genome-wide association (GWAS) data in the prospective CANTO cohort.

Abstract No: 11515

**Background:** Many BC survivors report fatigue. The relevant genomic correlates of fatigue after BC are not well understood. We applied a previously validated machine learning methodology (Oh 2017) to GWAS data to identify biological correlates of fatigue induced after tx.

**Methods:** We analyzed 3825 BC pts with GWAS data (Illumina InfiniumExome24 v 1.1) from the CANTO study (NCT01993498). The outcome of this study was post-tx fatigue 1 year after the end of primary chemotherapy/ radiotherapy/ surgery using the EORTC C30 fatigue subscale (overall fatigue) and the EORTC FA 12 fatigue domains (physical/emotional/cognitive). For each domain, we limited the study group to those with zero baseline fatigue and defined severe fatigue change as score increase above the third quartile. We tested univariate correlations between severe fatigue in each domain and 496539 SNPs as well as relevant clinical variables. The machine learning prediction model based on preconditioning random forest regression (PRFR) (Oh et al., 2017), was then built using the SNPs with ancestry adjusted univariate p-value < 0.001 and clinical variables with Bonferroni adjusted p-value < 0.05. The model was validated in a holdout subset of the cohort. Gene set enrichment analysis (GSEA) was performed using MetaCore to identify key biological correlates relevant to tx-induced fatigue.

**Results:** Distinct results were found by fatigue domain (table). GSEA showed that the cognitive fatigue model SNPs included biomarkers for cognitive disorders ($p = 1.6 \times 10^{-12}$) and glutamatergic synaptic transmission ($p = 1.6 \times 10^{-8}$).

**Conclusions:** A SNP based model had differential performance by fatigue domain, with a potential genetic role on risk and biology for tx induced cognitive fatigue. Further research to explore biomarkers of tx induced fatigue are needed.

# Patient Outcomes

## Prediction of hepatocellular carcinoma patient survival using machine learning classification rules.

Abstract No: e15649

**Background:** The outcome prediction of hepatocellular carcinoma (HCC) is conventionally determined by evaluating tissue samples obtained during surgical removal of the primary tumor focusing on their clinical and pathologic features. Recently, accumulating evidence suggests that cancer development is comprehensively modulated by the host's immune system underlying the importance of immunological biomarkers for the prediction of HCC prognosis. However, an integrated predictive algorithm, incorporating clinical characteristic and immune features, remains to be established.

**Methods:** We obtained respectable stage II HCC specimens, along with adjacent para-tumor tissues from 221 patients who underwent surgical resection at Eastern Hepatobiliary Surgery Hospital, (Shanghai, China) from 2015 through April 2018. Characteristics such as CD8+, CD163+, tumor-infiltrating lymphocytes (TILs) were obtained for further model construction used to predict the status of 3 survival indexes: Overall Survival (OS ,≤ 24 or > 24 month), Progression Free Survival (PFS, ≤ 6 or > 6 month), and Recurrence/Death (RD). Mutual information and coefficient between each feature and the survival indexes were tested to remove low scoring features after data cleaning and standardization. Furthermore, recursive features selection was preformed to obtain the optimal features combination. Finally, supervised learning techniques include either boosting or bagging strategy were used to fit and predict model with a grid-search method optimizing the parameters. Meanwhile, a cross validation procedure with 0.2 proportion of test cohort was randomly carried out for 10 times to evaluate the model.

**Results:** We finally confirmed 15 biomarkers from the 46 candidates as features for the survival status prediction by using a 221-patient cohort. Among them, the top 10 most important biomarkers, included both clinical and immune attributes. The AUC of our model for survival indexes (OS, PFS, RD) was ranged from 0.76 (RD) to 0.8 (PFS), and the accuracy was above 0.85.

**Conclusions:** We describe the integrative analysis of the clinical and immune features which collectively contribute to the survival index of HCC. Machine learning techniques, such as Gradient Boosting and random forest classifier, have a great promise for using in HCC cancer survival prediction.

# Development of an artificial intelligence model to predict survival at specific time intervals for lung cancer patients.

Abstract No: 6556

**Background:** Survival prediction models for lung cancer patients could help guide their care and therapy decisions. The objectives of this study were to predict probability of survival beyond 90, 180 and 360 days from any point in a lung cancer patient's journey.

**Methods:** We developed a Gradient Boosting model (XGBoost) using data from 55k lung cancer patients in the ASCO CancerLinQ database that used 3958 unique variables including Dx and Rx codes, biomarkers, surgeries and lab tests from ≤1 year prior to the prediction point, which was chosen at random for each patient. We used 40% data for training, 25% for hyper-parameter tuning, 20% for testing and 15% for holdout validation. Death date available in the Electronic Health Record was cross checked by linkage to death registries.

**Results:** The model was validated on the holdout set of 8,468 patients. The Area Under the Curve (AUC) for the model was 0.79. The precision and recall for predicting survival beyond the three time points were between 0.7-0.8 and 0.8-0.9 respectively (see table). This compares favorably to other lung cancer survival models created using different machine learning techniques (Jochems 2017, Dekker 2009). A Cox-PH model created using the top 20 variables also had a significantly lower performance (see table). Analysis of input variables yielded distinctive patterns for patient subgroups and time points. Tumor status, medications, lab values and functional status were found to be significant in patient sub cohorts.

**Conclusions:** An AI model to predict survival of lung cancer patients built using a large real-world dataset yielded high accuracy. This general model can further be used to predict survival of sub cohorts stratified by variables such as stage or various treatment effects. Such a model could be useful for assessing patient risk and treatment options, evaluating cost and quality of care or determining clinical trial eligibility.

# Using machine learning to predict mortality in older patients with cancer: Decision tree and random forest analyses from the ELCAPA and ONCODAGE prospective cohorts.

Abstract No: 11516

**Background:** Accurate prognosis is crucial to decision making in oncology but remains challenging in older patients due to the heterogeneity of this population and the lack of ability of current models to capture complex interactions between oncological and geriatric predictors. We aimed to develop new predictive algorithms based on machine learning to refine individualized prognosis in older patients with cancer.

RAS

**Methods:** Data were collected from 3409 patients ≥70 years referred to geriatric oncology clinics for completion of a geriatric assessment (GA), including 2012 and 1397 patients from the ELCAPA (training set) and ONCODAGE (validation set) French prospective cohorts, respectively. Candidate predictors included baseline oncological and geriatric parameters, G-8 score and routine biological data (CRP/albumin ratio). Prognostic models for 12-months mortality were built using Cox regression model, single decision tree (DT) and random survival forest (RSF). Models performance was compared based on externally validated Harrell's C-indexes.

**Results:** During the 1-year study period, 875 (43%) and 219 (16%) patients died in the training and validation sets, respectively (mean age: 81±6 / 78±5, women 47% / 70%, metastasis 50% / 34%). Cox model identified 9 independent predictors of mortality: tumor site/metastatic status, anticancer treatment, weight loss > 3kg, drugs > 5, renal failure, increased CRP/Albumin, ECOG-PS≥2, ADL≤5 and altered TGUG. DT identified more complex combinations between features, yielding 16 patient groups with highly differentiated survival, notably depending on the G-8 (< 10 vs. ≥10 as the root node). RFS had the highest C-index (0.86 [RFS], 0.82 [Cox], 0.81 [DT]), identifying the G-8, CRP/albumin and tumor site/metastasis as the most important features.

**Conclusions:** While Cox modeling confirmed known independent prognostic factors, DT revealed more complex interactions between them, and random forest achieved superior prognostic performance by better capturing patient's complexity. The latter model has been implemented into an interactive web interface for easy and direct use in clinical practice. Clinical trial information: NCT02884375

## ml-RECIST: Machine learning to estimate RECIST-defined outcomes in NSCLC patients treated with PD-(L)1 blockade.

Abstract No: 9052

**Background:** Real-world evidence (RWE) is increasingly important for discovery and may be an opportunity for regulatory approval. Effective use of RWE relies on determining treatment-specific outcomes, such as overall response rate (ORR) and progression-free survival (PFS), which are challenging to accurately evaluate retrospectively and at scale. We hypothesized the use of machine learning of text radiology reports from patients with NSCLC treated with PD-1 blockade could be used to train a model that estimates RECIST-defined outcomes.

**Methods:** 2753 imaging reports from 453 patients with advanced NSCLC treated with PD-1 blockade were collected and separated into independent training (80%, n = 362) and validation (20%, n = 92) cohorts. Reports were limited to interval of PD-1 blockade. RECIST reads performed by thoracic radiologists on all patients served as "gold standard" to determine ORR, occurrence of, and date of progression. Baseline reports were compared to all follow up reports to determine machine-learning RECIST (ml-RECIST). A four layers neural-network model for classification was proposed to solve the three above tasks.

**Results:** In the training cohort, ml-RECIST best estimated ORR by RECIST (accuracy CR/PR 84%, SD 82%, POD 91%). ml-RECIST estimated PFS by RECIST accurately predicting progression occurred at any time (86%) and exact progression date (65%). Date of progression was closely correlated (Pearson's r coefficient = 0.91, 95% CI:0.89-0.94, p < 0.001) in patients determined to have progressed by both methods. Similar accuracy of ml-RECIST was observed in the validation cohort (accuracy CR/PR 84%, SD 80%, POD 90%; progression occurred 86%, progression date 72%). Accuracy was consistent when RECIST reads were performed prospectively as part of clinical trials vs retrospectively for standard of care treatment (e.g. CR/PR 82% vs 88%, respectively). ml-RECIST-defined response similarly determined improvement in overall survival compared to RECIST (HR = 0.19, p < 0.001 vs HR = 0.26, p < 0.001 respectively).

**Conclusions:** Machine learning-RECIST ("ml-RECIST") accurately estimates outcomes using imaging text reports. ml-RECIST may be tool to determine outcomes expeditiously and at scale for use in RWE studies, enabling more useful and reliable applications of large clinical databases.

# Patient Engagement
## (Adherence, Satisfaction and Co-Morbidity Management)

## Leveraging a conversational agent to support adherence to oral anticancer agents: A usability study.

Abstract No: 6534

**Background:** Identifying effective, scalable strategies to ensure patient adherence to oral anticancer agents (OACAs) is a major challenge. Previous studies have shown widely variable rates of adherence, and suboptimal adherence is associated with decreased effectiveness and higher costs. A small but growing literature supports digital health

behavioral interventions across a variety of chronic illnesses, including in cancer. In particular, conversational agents— or technologies that mimic human conversation using text or spoken language—have shown early promise in supporting behavior change but have yet to be rigorously tested in the context of OACAs.

**Methods:** A rapid cycle prototyping approach led to the development of 'Penny'– a bidirectional, conversational agent that engages patients via text messaging, and leverages natural language processing and machine learning to learn from clinical interactions. Core functionalities include: (1) real-time dosing instructions, (2) motivational reminders, and (3) symptom monitoring with self-management support. We conducted a four-month usability study between December 24, 2017 and May 1, 2018 in a large academic cancer center. At monthly intervals for the first 12 weeks of follow-up, research staff conducted qualitative interviews with participants to evaluate usability and acceptance.

**Results:** 11 patients with gastrointestinal neuroendocrine cancer on capecitabine and temozolomide were approached regarding the study. Of these, 10 agreed to participate (ages 45 to 71). Overall, participant satisfaction was high with a Net Promoter Score of 100. Reliability of Penny's algorithmic branching to provide accurate dosing information and symptom triage was also high: symptoms were accurately graded 100% of the time, and there was appropriate self-management advice or provider triage 100% of the time. Average daily adherence (based on self-report) was 98%. Participants reported that 3 emergency room visits were avoided during the study period.

**Conclusions:** In preliminary testing, a mobile phone-based conversational agent was a usable and acceptable means of supporting OACA adherence. Expanded study testing patient safety and efficacy is underway.

# Artificial intelligence-based clinical decision-support system improves cancer treatment and patient satisfaction.

Abstract No: e18303

**Background:** Traditional diagnostic model for cancer heavily relies on physicians and their teams' knowledge. However, under this diagnostic model, patients' source of information is quite limited. Cancer patients usually fill with negative emotion. Lack of knowledge to the disease and treatment options further leads to less confidence to their treatment outcome.

**Methods:** In order to improve their faith in getting proper treatment and the hope for surviving the deadly disease, we have introduced an artificial intelligence based clinical decision-support system, the Watson for Oncology (WFO), since May-2018. WFO is developed by IBM, it assesses information from a patient's medical record, evaluates medical evidence, and displays potential treatment options. Our oncologist can then apply their own expertise to identify the most appropriate treatment options. We have generated a new 7-step consultation system with the help of WFO. That include 1: introduce the WFO to patients, 2: patients express their demands and expectations, 3: the oncologist presents patient's medical condition, 4: discussion with other members in the consultation team, 5: input patients' information into WFO system and review treatment options, 6: discuss and finalize treatment options with patients, 7: feedbacks form patients after consultation. 70 patients who were hospitalized from May-2018 to Dec-2018 were divided into two groups, 50 patients volunteered to be assigned to the new 7-step consultation system and 20 patients stayed with the traditional diagnostic method to find them treatment options. All patients were followed up by questionnaire.

**Results:** The results showed that patients in the 7-step consultation group presented significantly higher satisfaction rate towards treatment options, confidence level to their health care workers, and willingness to follow the treatment option when compared to patients in the traditional diagnostic group.

**Conclusions:** The WFO assisted 7-step consultation system not only provides a more efficient way to find treatment options, but also improves patients' understanding to their disease and possible side effects towards the treatment. Most importantly, patients build stronger confidence with their health care team and are willing to believe they will benefit from the treatment plans.

# A randomized controlled trial of a novel artificial intelligence-based smartphone application to optimize the management of cancer-related pain.

Abstract No: 11514

**Background:** Cancer pain is a significant problem that impairs patient quality of life and increases healthcare utilization. ePAL is a smartphone application that utilizes patient-reported outcomes (PROs) and artificial intelligence (AI) to optimize cancer pain management. This randomized controlled trial examined the impact of ePAL on cancer pain severity, attitudes toward cancer pain, and healthcare utilization.

**Methods:** Patients with pain from metastatic solid tumors (n = 112) undergoing treatment in a palliative care clinic were randomized to either a control group (n = 56) that received usual care or an intervention group (n = 56) that received ePAL in addition to usual care for 8 weeks. Measures of pain severity (Brief Pain Inventory), attitudes towards

cancer treatment (Barriers Questionnaire II) and anxiety (General Anxiety Disorder-7) were assessed. We used repeated measures mixed modeling to assess change in outcome measures over time. We also conducted a chart review to identify pain-related hospital admissions and emergency department (ED) visits and compared risk between study groups.

**Results:** Pain severity (BPI) and negative attitudes toward cancer treatment (BQ-II) decreased significantly for those assigned to ePAL compared to controls (ß = -0.09, p = 0.034 and ß = -0.037, p = 0.042, respectively). Patients assigned to ePAL reported higher anxiety scores compared to controls (ß = 0.21, p = 0.015). Patients assigned to ePAL had significantly fewer pain-related hospital admissions (n = 4 vs. n = 20, per patient risk ratio 0.31, p = 0.018) and fewer pain-related admissions through the ED (n = 2 vs. n = 14, per patient risk ratio 0.18, p = 0.008) compared to control group.

**Conclusions:** To our knowledge, this is the first mobile app to utilize patient reported outcomes and artificial intelligence to significantly decrease pain scores and pain-related hospitalizations in patients with cancer-related pain. Future directions include examining the efficacy of ePAL in settings with limited access to palliative care.