Evaluation of deep learning techniques (DL) in RNA sequencing data for the prediction of metastatic renal cell cancer m(RCC) patient response to immune checkpoint inhibitors.

Authors: Sandra Alonso, Ignacio Duran, Jesús García-Donas, Cristina Rodríguez, Enrique Grande, Juan Diego Cacho, Laura Rodríguez, Patricia Saiz, Guillermo Crespo, Laura Villalobos, Carmen Perezagua, Lucía González, Olatz Etxaniz, Álvaro Pinto, Alejo Rodríguez-Vida, Pablo Borrega, Natalia Vidal, Nuria Romero, José Luis Ayala.

Background:
Immune checkpoint inhibitors have become a cornerstone in the management of mRCC. However, to identify the most suitable patients for this treatment is an unmet medical need. We aimed to explore the utility of DL integrating clinical and molecular data to predict response to immunotherapy.

Methods: We conducted a retrospective analysis using publicly available data from patients treated with nivolumab in the clinical trials Chekmate 009, 010 and 025. The primary objective was to assess the performance of different DL models (autoencoder and convolutional neural network (CNN)) predicting the PFS of these patients.
With that scope, we followed several research lines including the creation of combined datasets with clinic and RNA sequencing and comparing the results of the DL models against the performance of traditional machine learning (ML) models. Finally, we came up with an interpretability analysis of those black-box models using LIME and SHAP values.

Results: Clinical and transcriptomic data were available from 181 nivolumab-treated patients. Outcomes achieved confirmed that we can model the response for NIVOLUMAB using RNA sequencing data. [Table 1] However, DL models have not demonstrated to be significantly better than traditional ML methods when predicting response (p= 0.068). Deep autoencoder provided 68.9% accuracy, but the most accurate model was logistic regression classifier which achieved 86.4% of hit rate. Interpretability results revealed that most relevant genes for the decision making were related with the immune response and the regulation of kinases. Regarding interpretability, best results were achieved integrating both transcriptomic and clinical data (7 out of 10 DL and ML tested models achieved higher hit rates with combined data set. Logistic regression classifier improved its accuracy in 11%).

Conclusions: The integration of clinical and molecular data could lead to more accurate predictions of outcome than any dataset by its own. However, further research is intended in the field of the DL analysis, as data codification and data structure could bias the results. The ongoing study **ART** (**A**rtificial Intelligence in **R**enal **T**umors) will address this issue prospectively.

Model	**Data set**	**Accuracy**
Logistic regression	30 most reliable genes and MSKCC variable	0.846
Deep autoencoder	30 most reliable genes and 12 clinic variables	0.689
CNN	400 most reliable genes	0.675

Table 1: Results in terms of accuracy for DL and ML best scored models.