

# Biomedical Impact Of Lung Cancer Knowledge Graph Part 3

Case Study

## Introduction

Discovering new effective therapies is an almost impossible feat, with only 5 in 5,000 new therapies (0.1%) in the pre-clinical phase entering clinical trials. In addition to the high drop off rate of therapies, there continues to be a high unmet medical need for novel therapies, especially in complex and hard to treat areas such as lung cancer. Galactic AI™ provides a unique avenue for exploring potential drug targets that may increase the chance of successfully developing therapies. Galactic AI™ carefully curates data to generate knowledge graphs, which documents protein interactions in the context of the disease. In this case, the lung cancer knowledge graph consists of 71K distinct directed molecular interactions and 7.8K distinct proteins, all documented in the context of lung cancer research. These interactions can be analysed and used to uncover potential drug targets.

## Solution

The initial steps to seek out novel lung cancer targets involve the creation of the lung cancer knowledge graph (see knowledge graph case study). Existing lung cancer drug targets are overlaid on the knowledge graph and tested using various methods (see Analysis causal data and knowledge graphs case study) to reveal potential targets. The scoring and ranking methodologies used to identify potential protein drug targets, of which 11 have an exceptionally high score - 6 are already known targets and 3 are targets for other indications.

The contextual nature of the knowledge graph can be taken advantage of to explore different lung cancer diseases, for example non-small cell lung cancer (NSCLC) vs. small cell lung cancer (SCLC), two major types of lung cancer. The comparative analysis reveals 5K interactions and 1.7K genes for NSCLC and 32.9K interactions

and 5.2K genes for SCLC. Drilling down into disease subtypes provides a more granular investigation of a disease area and easy comparison while also providing context to disease subtypes.

While we can use the lung cancer knowledge graph to predict potentially viable novel lung cancer targets how can we be confident in its accuracy? To determine the accuracy, we create a blind test by looking retrospectively at the data from a certain time point and testing if our algorithm can correctly predict drug targets that have yet to be discovered. All the data published from 2010 onwards is removed from the knowledge graph and we try to predict new drug targets that emerged from 2011 onwards. The filtering of the knowledge graph excluded 55.1K protein interactions published after 2010, leaving 15.7K interactions. Target scores were recalculated with the filtered data and measured against scores derived from the full graph. We show that targets of drugs approved after 2011 are significantly more likely to be scored higher ( $p=0.08$  Mann-Whitney U Test). This reinforces the viability of the novel lung cancer target prediction algorithm. The list of possible targets that are scored highly are ready for further investigation.

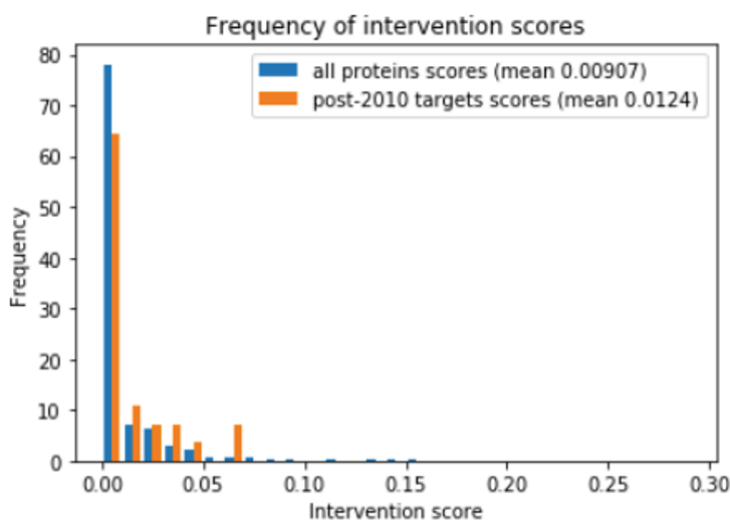


Figure 1: Frequency of intervention scores before and after 2010

Another important area to investigate is the specificity of the predicted lung cancer targets. One way to measure specificity is to measure target likeness between general drug targets and lung cancer targets. To measure this, we devised one support vector machine (SVM) for lung cancer targets and one SVM for general targets, using the protein similarity scoring algorithm for upstream genes supplied by Galactic AI™. Plotting these scores, shows how certain targets have high lung cancer target likeness, but also high general target likeness (low specificity). High specificity lung cancer targets are those that have high lung cancer target likeness and low general target likeness.

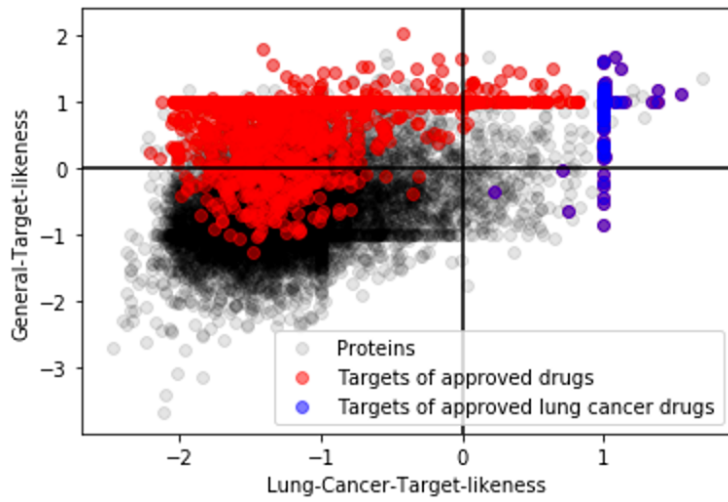


Figure 2: Lung cancer Target Likeness Versus General Drug Target Likeness

Furthermore, overlaying our lung cancer target prediction scores reveals targets that have both high lung cancer specificity and fit the profile of a successful lung cancer drug target (the darker blue dots in Figure 3). These hypotheses can be explored generally or more specifically across different lung cancer diseases.

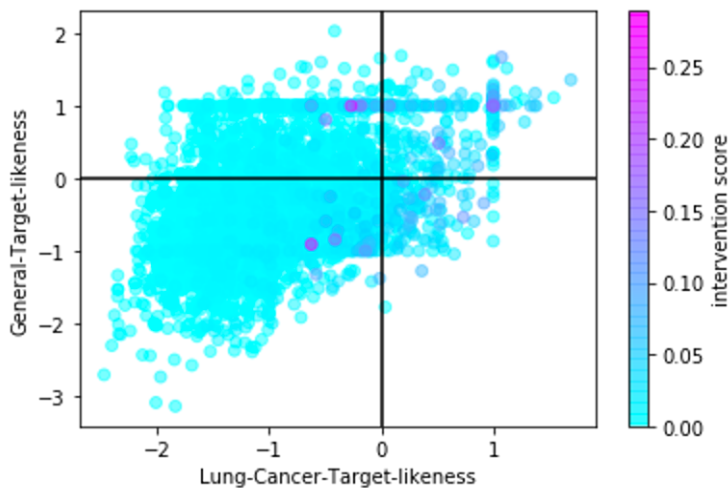


Figure 3: Drug Target likeness Shown Against the Lung Cancer Intervention Score

Furthermore, overlaying our lung cancer target prediction scores reveals targets that have both high lung cancer specificity and fit the profile of a successful lung cancer drug target (the darker blue dots in Figure 3). These hypotheses can be explored generally or more specifically across different lung cancer diseases.

## Conclusion

Using Galactic AI™ technology we can automatically and quickly build knowledge graphs under any selected criteria of interest. Galactic AI™ has shown its ability to not only correctly identify drug targets with a high average precision, but also accurately hypothesise novel predicted targets. In addition, using Galactic AI™ we

can illustrate how lung cancer targets can be stratified across different disease subtypes and ranked by their specificity. These case studies have focused on identifying novel drug targets in lung cancer, however the unique Galactic AI™ technology and analysis can be used across any disease.

## Impact / Benefits

- **Advancing science** –Galactic AI™ allows exploration of a disease area in granular detail and provides the potential to uncover key findings that can direct future research efforts and reinforce existing hypotheses
- **Unique** - Galactic AI™ captures contextual causal data not available in any other database and enables complex algorithms to make novel and accurate predictions on drug targets
- **Multiple applications** – As well as determining targets, Galactic AI™ can also be used to predict biomarkers, drug-drug interactions, off-target effects, likely safety and toxicology issues and many other systems biology applications

