1. **Understanding data and background**

<https://www.kaggle.com/competitions/lish-moa/discussion/184005>

We have been given data collected from a cell-based assay which records levels of gene expression and cell viability. Our objective is to use the patterns in this data to predict the mechanism of action (MoA) of a drug. This is a multi-label competition where some rows can be linked to multiple targets, instead of multi-class. The data comprises of 772 gene expression features labeled as "g-" that represent the expression of one specific gene, and 100 cell viability features labeled as "c-" that represent the viability of one particular cell line. The original data was normalized using [quantile normalization](https://www.youtube.com/watch?v=ecjN6Xpv6SE&ab_channel=StatQuestwithJoshStarmer).

![Chart, bar chart

Description automatically generated]()

![Chart, bar chart

Description automatically generated]()

![Chart, bar chart

Description automatically generated]()