***Variable Codebook***

The codebook is for the all\_data data frame and all\_data.csv file which includes both imported and calculated data.

**Factors**:

**formulation\_id**: An arbitrary, specific number ranging from 1-45 denoting the specific combination of lipid components tested for an experiment. This is not used for analysis but rather for data wrangling to assemble a full data frame from all data collected for the experiment. It is numeric but categorical and nominal in nature as it has no inherent quantitative meaning.

**DSPC**: The molar fraction of the formulation component DSPC out of the total moles of lipid. It is quantitative and continuous. The levels were preselected using a bounded D-Optimal design for all formulation components and are given as 0, 0.1, 0.2, 0.0998, and 0.085.

**DMG.PEG**: The molar fraction of the formulation component DMG-PEG out of the total moles of lipid. It is quantitative and continuous. The levels were preselected using a bounded D-Optimal design for all formulation components and are given as 0.03, 0.065, 0.0668, and 0.1.

**Chol**: The molar fraction of the formulation component cholesterol out of the total moles of lipid. It is quantitative and continuous. The levels were preselected using a bounded D-Optimal design for all formulation components as are given as 0, 0.085, 0.17, 0.227, 0.23, 0.25, 0.255, 0.26, 0.302, and 0.5.

**DOTAP**: The molar fraction of the formulation component DOTAP out of the total moles of lipid. It is quantitative and continuous. The levels were preselected using a bounded D-Optimal design for all formulation components as are given as 0, 0.1, 0.135, 0.17, 0.185, 0.2, 0.234, 0,235, 0.26, 0.27, 0.303, 0.325, 0.35, 0.36, 0.385, 0.4, 0.435, 0.45, 0.47, 0.485, 0.52, 0.64, 0.7, 0.77, and 0.8.

**MC3**: The molar fraction of the formulation component MC3 out of the total moles of lipid. It is quantitative and continuous. The levels were preselected using a bounded D-Optimal design for all formulation components as are given as 0, 0.1, 0.17, 0.185, 0.2, 0,234, 0.235, 0.26, 0,27, 0.303, 0.325, 0.35, 0.36, 0,385, 0.4, 0.435, 0.45, 0.47, 0.485, 0.514, 0.64, 0.668, 0.7, 0.735, 0.77, 0.8.

**cell\_type**: The tissue from which the cells were derived that the formulation was tested on. It is categorical and nominal. It can be one of four values: Liver, Lung, Muscle, or Immune.

**nucleic\_acid**: The type of nucleic acid tested in the observation. It is categorical and nominal. It can be one of two values: mRNA or DNA.

**plate\_number**: An arbitrary number assigned to each duplicate of experimental conditions tested. Plate 2 was prepared directly after plate 1 with the same formulation and the cells for both plates were derived from the same starter culture. It is categorical in nature and ordinal with respect to time.

**incorporation**: The nucleic acid formulation incorporation method used. It is categorical and nominal. The values can either be Encapsulated (nucleic acid mixed with formulation components during particle synthesis) or Adsorbed (nucleic acid mixed with nanoparticles following synthesis and buffer exchange.).

**dose**: The dose of nucleic acid added to each well of cells at the start of the transfection. It is categorical and ordinal. It can have the values of High (100 ng nucleic acid/well), Medium (50 ng nucleic acid/well), or Low (25 ng nucleic acid/well). It can be converted to a quantitative variable.

**time**: The time point the transfection quantification was made. It is quantitative and continuous. The cells were analyzed every 6 hours for 72 hours following treatment with the formulations.

**Responses**:

**tfx\_response**: The total fluorescent signal in each well following treatment under the parameters described in the other dataset variables, transformed logarithmically (base 10). It was computed by the IncuCyte built in image analysis software and given on a scale of arbitrary intensity units. It is quantitative and continuous on the range of 0.0 to 7.22.

**size**: The size of each formulation (unique only to the formulation\_id and nucleic acid) of the nanoparticles in each formulation in nm. It is quantitative and continuous. Values lie on the range 0 nm to 214.2 nm.

**efficiency**: The efficiency of loading the nanoparticles in the formulation (unique only to the formulation\_id and nucleic acid) with the nucleic acid therapeutic. It is on the range 0 to 99.0 %.

**Calculated Values**:

**initial\_response**: The tfx\_response value at time = 0 hrs for each transfection observation formatted into a column to facilitate signal to noise ratio calculations.

**average\_response\_twenty\_four**: The average tfx\_response for each transfection computed for all time points between 24 anf 72 hrs. It is formatted into a column to facilitate determination of responder status.

**signal\_to\_noise**: The signal to noise ratio for each tfx\_response as computed by the formula:

**responded**: A categorical variable that indicates the transfection responded to the formulation. A responder is defined as a transfection that produced an average signal to noise ratio of greater than 3 for the time range of 24 to 72 hours. It can have the value of No Response or Responded.

**quantifiable**: A categorical variable that indicates the transfection signal for this observation is quantifiable. A quantifiable signal was defined as one where the signal to noise ratio for that time point was greater than 10.