# Assignment 2

Erata Maharshi (210050049) Motupalli Yaswanth (210050099)

April 21, 2024

## Objective

This report presents the results of the analysis conducted to determine whether a localization sequence is present in specific genes. Our team has been assigned to find whether the given sequence **SKL(serine-lysine-leucine)** is a tripeptide sequence found at the C-terminus of proteins targeted to the peroxisomes within cells which serves as a peroxisome localization signal **PTS1**, is present within various genes. Below is a table summarizing the presence of this sequence within each gene.

## Code

We downloaded sequences for the given genes and saved them as text files. Then, we wrote a Python script to check whether the localization sequence is present in the gene. Finally, we saved the outputs in a excel file. You can run the script.py directly with all other files downloaded locally, or you can run the IPython notebook file in Google Colab.

## Results

| Gene | Localization Sequence Present |
|---|---|
| ALDH18A1 | YES |
| ALG12 | NO |
| ALG9 | YES |
| ALMS1 | YES |
| AMER1 | NO |
| AMMECR1 | NO |
| ANK1 | YES |
| ANKRD11 | YES |
| ANKRD26 | YES |
| ANKS3 | NO |
| ANKS6 | NO |
| ANO2 | NO |
| ANOS1 | NO |
| BUB3 | NO |
| C2CD3 | NO |
| CA4 | NO |
| CABCOCO1 | NO |

| Gene | Localization Sequence Present |
|---|---|
| CATIP | NO |
| CBY1 | NO |
| CC2D2A | YES |
| CCDC103 | NO |
| CCDC112 | NO |
| CCDC13 | YES |
| CCDC14 | NO |
| CEP83 | NO |
| CEP89 | YES |
| CEP97 | YES |
| CERKL | NO |
| CETN1 | NO |
| CETN2 | NO |
| CETN3 | NO |
| CFAP100 | NO |
| CFAP126 | NO |
| CFAP161 | NO |

## Observation

We observed that the localization sequence is present in the genes ALDH18A1, ALG9, ALMS1, ANK1, ANKRD11, ANKRD26, CC2D2A, CCDC13, CEP89, and CEP97.

## Reference

UniProt