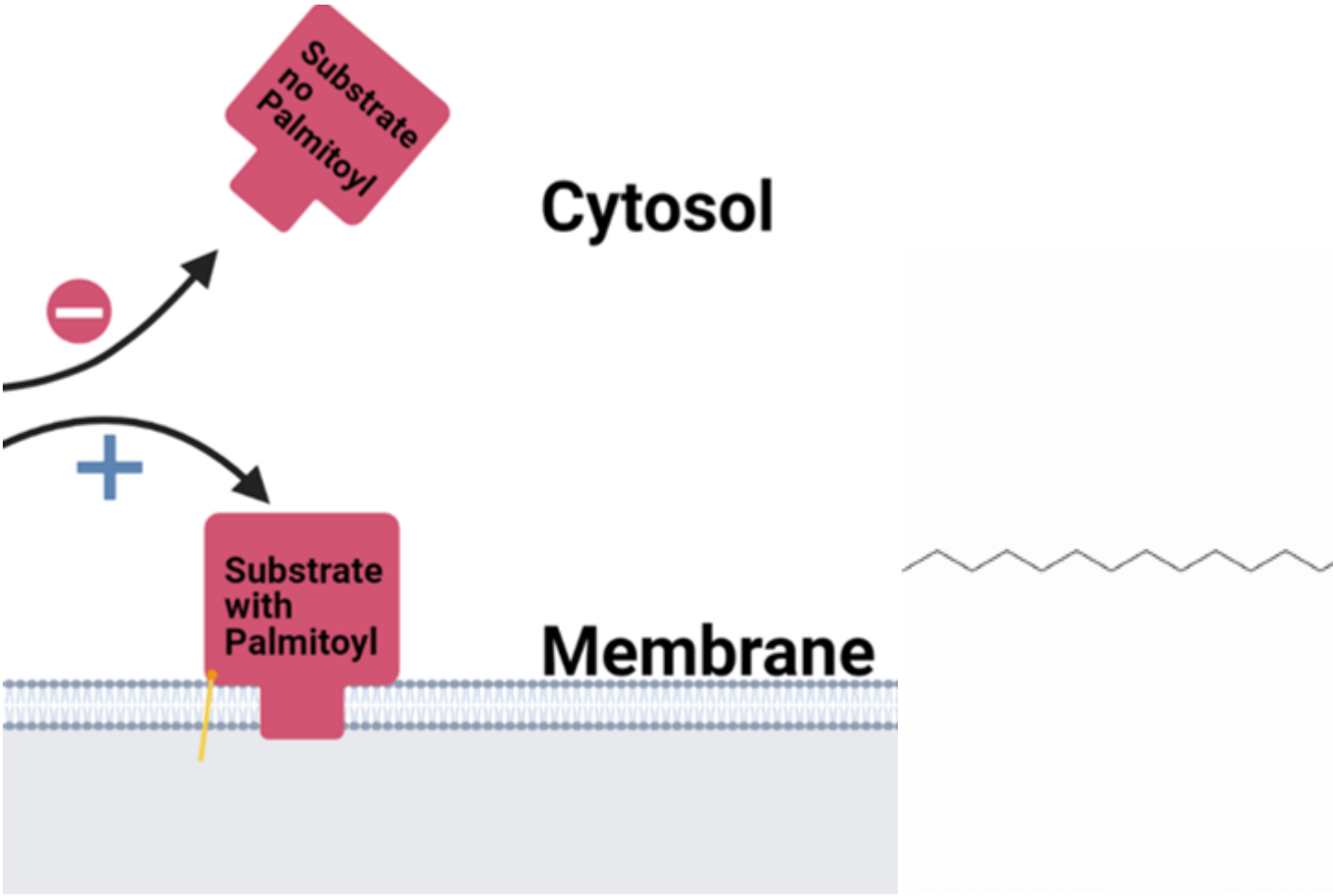
finalproj_re1032

Data analysis and Final Project for Data Science for Life Sciences (Fall2021).

https://github.com/robelliott4996/robelliott4996.github.io

Background of Palmitoylation

The data analyzed here comes from an acyl-RAC experiment performed on *Arabidopsis thaliana*, a plant model organism(Kumar, Carr, and Turner (2020)). Acyl-RAC, or acyl-resin assisted capture, is a protein purification method that captures S-acylated proteins. S-acylation, or palmitoylation, is a form of post-translational modification where a 16 carbon fatty acid, palmitate, is attached to a surface exposed serine residue. Because palmitate is significantly hydrophobic, this pushes the protein to favor attachment to a lipid membrane, with the palmitate group inserted into the lipid layer. This process is carried out by a family of enzymes called palmitoyl acyltransferases and their inverse enzyme, thioesterases.



Here, the authors identified 5262 unique palmitoylated proteins. Further, 12,419 modified cysteines were identified in those proteins.

Some Tables to Illustrate the Data

Below is a table of predicted subcellular locations and the number of genes from the dataset that are predicted to occupy them. These locations are determined using SUBAcon, the SUBcellular Arabidopsis consensus algorithm. At the bottom, some proteins are predicted to have more than one location; though the significant majority of them occupy a single location in the cell. Nine of these proteins have no prediction noted. I hesitate to assign strong support to these algorithm defined locations; if palmitoylation commonly changes the subcellular location of a protein then a single predicted location misses an extremely large depth of information.

Table 1: Table of Subcellular Protein locations as determined by SUBAcon

Subcellular Location	Proteins Counts
plastid	804
nucleus	854
extracellular	443
golgi	257
peroxisome	123
cytosol	1315
nucleus,cytosol	107
endoplasmic reticulum	201
mitochondrion	386
plasma membrane	528
vacuole	104
endoplasmic reticulum,golgi	10
mitochondrion,nucleus	2
mitochondrion,plastid	15
vacuole,golgi	6
plasma membrane,cytosol	21
endoplasmic reticulum,plastid	1
extracellular,plasma membrane	4
golgi,nucleus	1
endoplasmic reticulum,plasma membrane	12
endoplasmic reticulum,cytosol	2
NA	9
golgi,plasma membrane	9
golgi,cytosol	8
vacuole,plasma membrane	7
mitochondrion,cytosol	3
nucleus,plasma membrane	5
endoplasmic reticulum,extracellular	4
vacuole,cytosol	3
peroxisome,plastid	2
plastid,cytosol	3
extracellular,cytosol	2
mitochondrion,peroxisome	1
extracellular,plastid	1
plasma membrane,golgi,cytosol	2
nucleus,plastid,cytosol	2

Subcellular Location	Proteins Counts
vacuole,golgi,plasma membrane	1
peroxisome,cytosol	1
mitochondrion,plasma membrane	2
endoplasmic reticulum,vacuole	1

Next, proteins are organized according the MapMan bin they associated with. Similar to Gene Ontology terms, MapMan bins are a structured hierarchy of biological concepts originally developed specifically for *Arabidopsis*. Intensive studies have been done to compare the use of MapMan vs GO terms and found that it largely came down to a problem-specific condition.

The most interesting part is that of the 5262 proteins, 1678 of them are not assigned to any MapMan bin.

Table 2: Table of MapMan descriptors and Associated Protein Counts

MapMan description	Proteins counts
Lipid metabolism	173
not assigned	1678
Protein degradation	211
Cell wall	168
Phytohormones	117
Solute transport	166
Enzyme classification	311
Vesicle trafficking	181
RNA processing	146
Multi-process regulation	24
Cellular respiration	109
Coenzyme metabolism	118
Protein translocation	84
Amino acid metabolism	131
DNA damage response	12
Carbohydrate metabolism	129
Secondary metabolism	86
Protein biosynthesis	309
Redox homeostasis	48
Cell cycle	104
Protein modification	329
RNA biosynthesis	151
Nucleotide metabolism	46
External stimuli response	54
Photosynthesis	142
Cytoskeleton	117
Chromatin organisation	84
Polyamine metabolism	7
Nutrient uptake	27

Heading to the Bar (graphs)

The above graph illustrates an important aspect of palmitoylation and membrane association. Transmembrane helices(TMH) are naturally occurring structures that favor insertion into lipid membranes and as such can become membrane associated without palmitoylation. The more TMH present in a protein, the

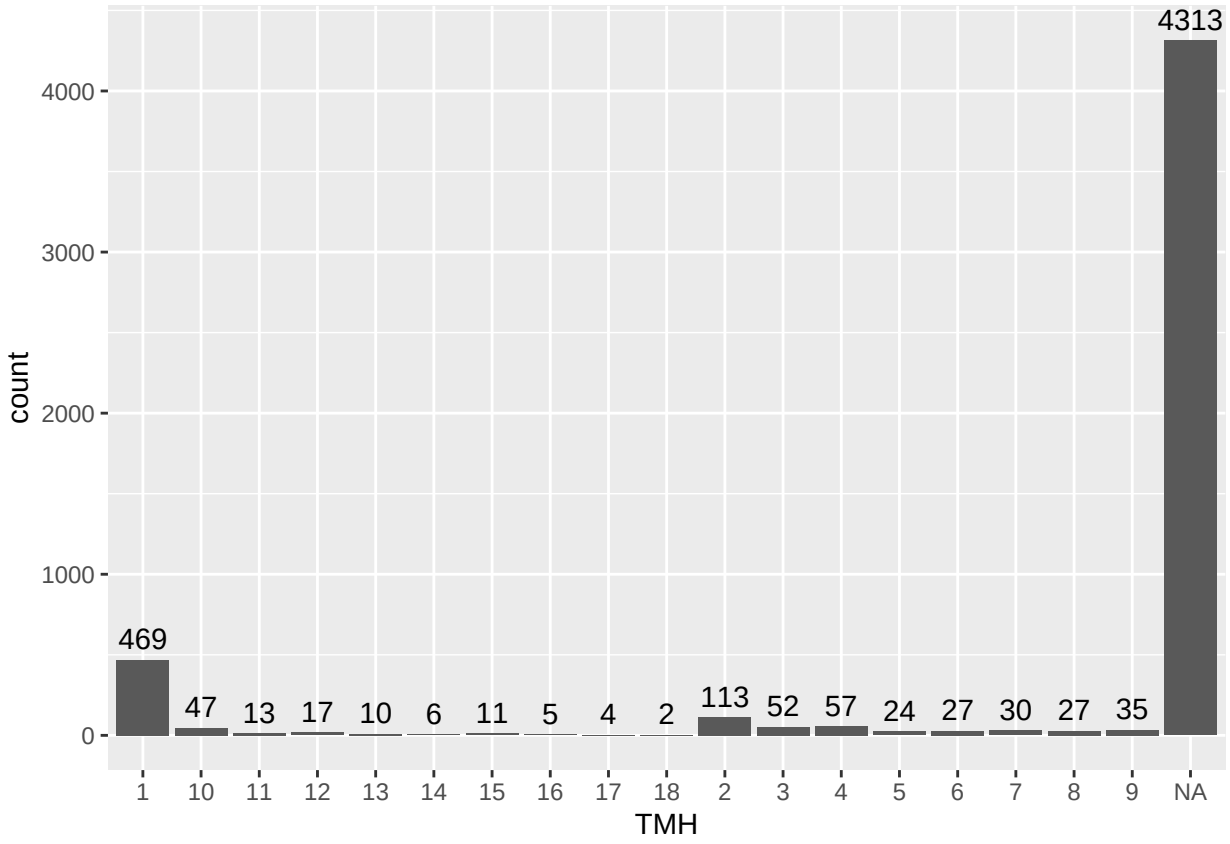


Figure 1: Number of Predicted Transmembrane Helices of Acylated Proteins

stronger the association to the point where some proteins do not assume their active conformation until they've inserted into a membrane. We see above that the overwhelming majority of proteins have no TMH, and would thus require palmitoylation in order to become membrane associated. Proteins with TMH would might require palmitoylation for some form of catalysis control.

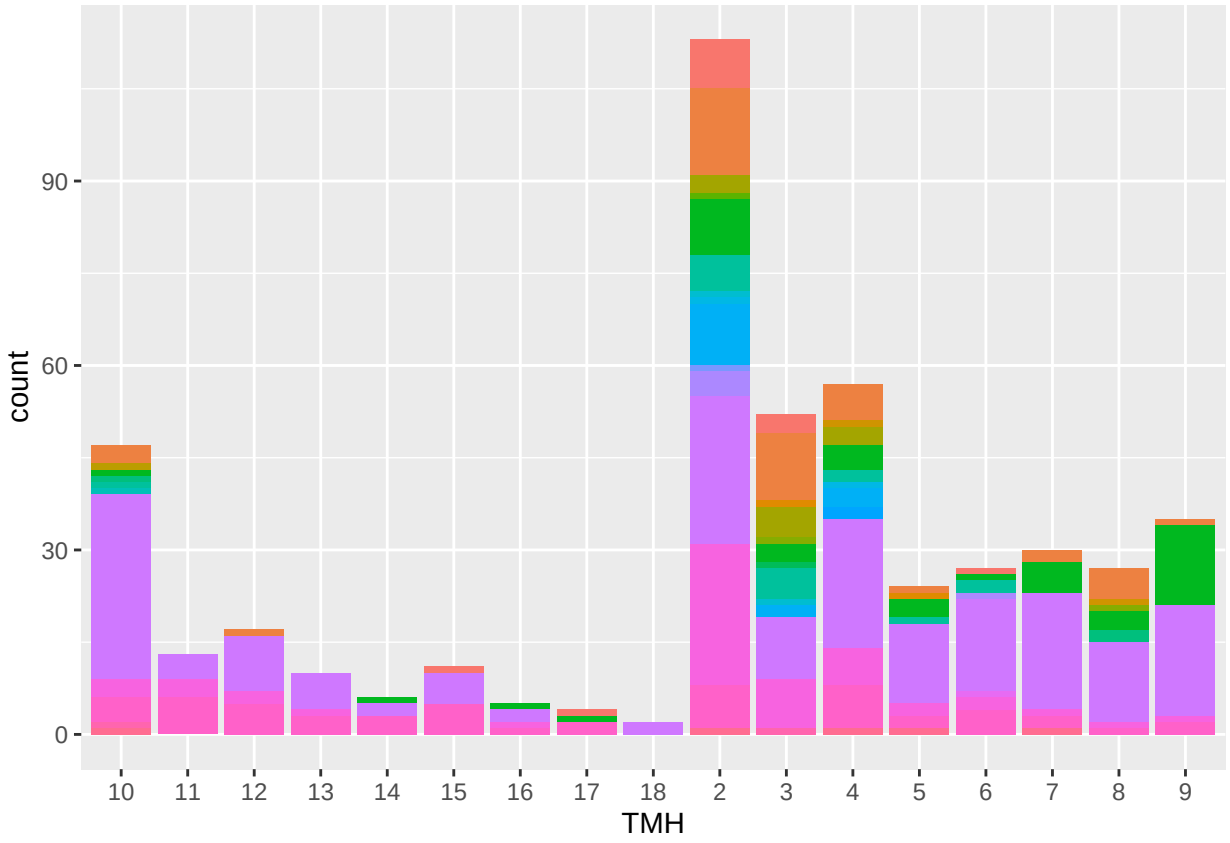
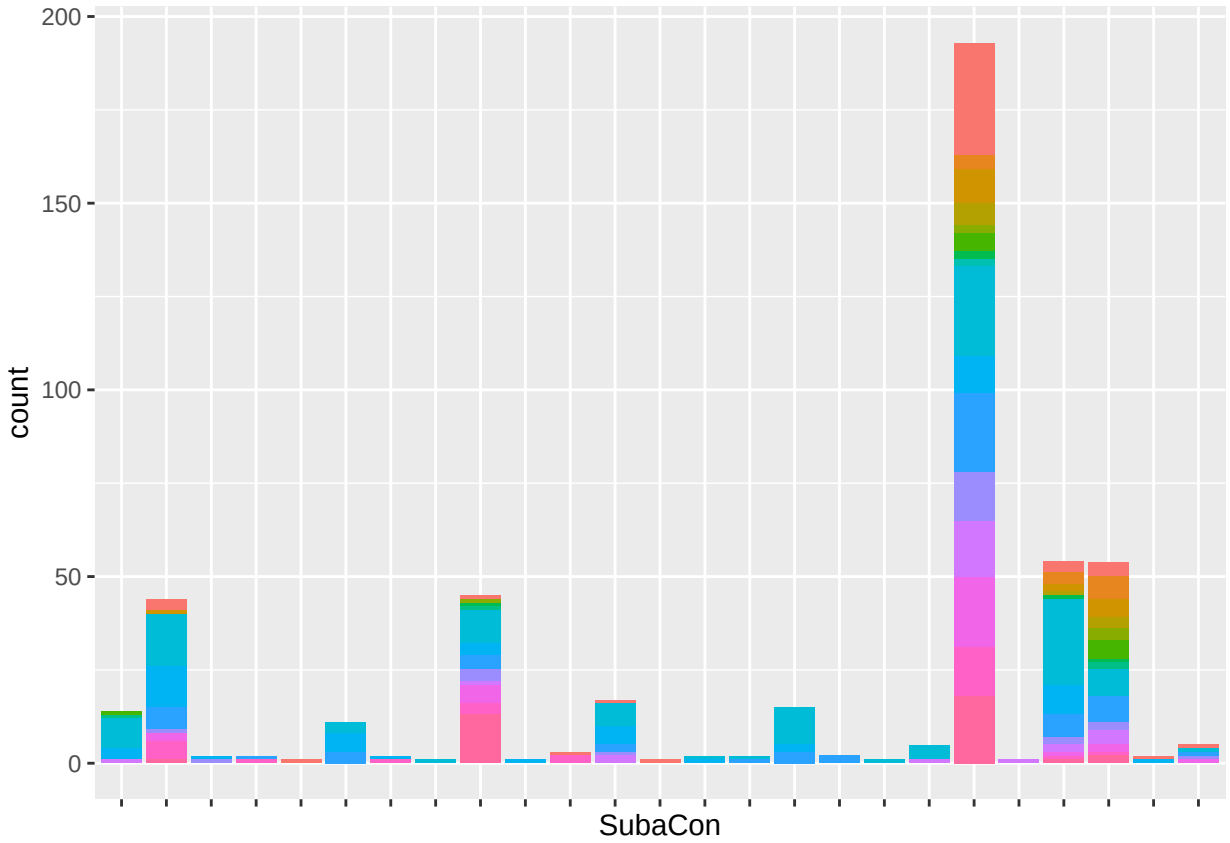


Figure 2: Number of Predicted Transmembrane Helices of Acylated Proteins excluding those with No or only 1 TMH; filled by subcellular location

However, when the proteins are separated by number of TMH and colored by subcellular location, the variation between conditions is negligible.



We see that above, when proteins are graphed according to subcellular location and then filled by number of TMH, that there is still an fairly even spread across all categories. If there is some overall defining characteristic of these proteins, it's not in their cellular location, or in the number of transmembrane helices.

References

Kumar, Manoj, Paul Carr, and Simon Turner. 2020. "An atlas of Arabidopsis protein S-Acylation reveals its widespread role in plant cell organisation of and function." *bioRxiv*. https://doi.org/10.1101/2020.05.12.090415.

Appendix