

**NOVEL SIGNALS IN MUTANT CALRETICULIN ENHANCE ITS SECRETION IN
MYELOPROLIFERATIVE NEOPLASMS**

A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF SCIENCE AND ENGINEERING OF
BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR MASTER OF SCIENCE IN
MOLECULAR BIOLOGY AND GENETICS

By:

SCOTT DAVID AUERBACH

AUGUST 2024

**NOVEL SIGNALS IN MUTANT CALRETICULIN ENHANCE ITS SECRETION IN
MYELOPROLIFERATIVE NEOPLASMS**

By Scott David Auerbach

August 2024

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Ilyas Chachoua (Advisor)

Serkan Belkaya

Ahmet Acar

Approved for the Graduate School of Science and Engineering:

Orhan Arıkan

Director of the Graduate School

ABSTRACT

NOVEL SIGNALS IN MUTANT CALRETICULIN ENHANCE ITS SECRETION IN MYELOPROLIFERATIVE NEOPLASMS

By Scott David Auerbach

M.Sc. in Molecular Biology and Genetics

Advisor: Ilyas Chachoua

August 2024

Calreticulin is an endoplasmic reticulum resident chaperone that is involved in many physiological roles including protein folding and calcium homeostasis. Recently, frameshift mutations have been documented in the exon 9 of the *calreticulin* (*CALR*) gene in myeloproliferative neoplasms. These mutations result in the substitution of the negatively charged sequence, in the carboxy-terminal domain, with positively charged and hydrophobic residues, which leads to the activation of the JAK/STAT pathway via direct interaction with the thrombopoietin receptor (MPL). Moreover, the loss of the endoplasmic retention signal (KDEL) facilitates its secretion to the extracellular medium, where it acts as a rogue cytokine. In this work, we aim to gain more understanding of the molecular mechanism behind calreticulin mutant secretion. To this end, we performed *in silico* studies where we analyzed the newly generated sequence of CALR mutants, and we show that this sequence is highly enriched in secretion signals containing mainly arginine. By using two online prediction tools we predict that mutating these signals into alanines significantly reduced the secretion of the mutants. Interestingly, introducing these signals to the wild type CALR enhances its secretion despite the presence of the KDEL signal. Altogether, these results challenge our current understanding of the molecular

mechanisms behind protein secretion and open new avenues towards the development of novel treatment strategies for CALR-associated MPNs by targeting the predicted signals with small molecules.

Keywords: Calreticulin, myeloproliferative neoplasms, *in silico* predictions, arginine-rich motifs, signals promoting extracellular secretion.



ÖZET

MUTANT KALRETİKÜLİN'DEKİ YENİ SİNYALLERİ MİYELOPROLİFERATİF NEOPLAZMLARDA SALGININI ARTTIRIR

Scott David Auerbach

Moleküler Biyoloji ve Genetik Yüksek Lisans

Tez danışmanı: Ilyas Chachoua

Ağustos 2024

Kalretikülin, protein katlanması ve kalsiyum homeostazisi dahil olmak üzere birçok fizyolojik rolde yer alan, endoplazmik retikulumda yerleşik bir şaperondur. Son zamanlarda, miyeloproliferatif neoplazmlarda *kalretikülin* (CALR) geninin ekson 9'unda çerçeve kayması mutasyonları belgelenmiştir. Bu mutasyonlar, karboksi terminal alanındaki negatif yüklü dizinin pozitif yüklü ve hidrofobik kalıntılarla değiştirilmesiyle sonuçlanır; bu, trombopoietin reseptörü (MPL) ile doğrudan etkileşim yoluyla JAK/STAT yolunun aktivasyonuna yol açar. Dahası, endoplazmik tutma sinyalinin (KDEL) kaybı, bunun hücre dışı ortama salgılanmasını kolaylaştırır ve burada serseri bir sitokin gibi davranır. Bu çalışmada kalretikülin mutant salgısının ardındaki moleküler mekanizmanın daha iyi anlaşılmasını amaçlıyoruz. Bu amaçla, CALR mutantlarının yeni üretilen dizisini analiz ettiğimiz *in silico* çalışmaları gerçekleştirdik ve bu dizilimin, esas olarak arginin içeren salgı sinyallerinde oldukça zengin olduğunu gösterdik. İki çevrimiçi tahmin aracını kullanarak, bu sinyalleri alaninlere dönüştürmenin, mutantların salgılanmasını önemli ölçüde azalttığını tahmin ediyoruz. İlginç bir şekilde, bu sinyallerin vahşi tip CALR'ye dahil edilmesi, KDEL sinyalinin varlığına rağmen onun salgılanmasını artırır. Toplamda bu sonuçlar, protein salgılanmasının ardındaki moleküler

mekanizmalara ilişkin mevcut anlayışımıza meydan okuyor ve tahmin edilen sinyalleri küçük moleküllerle hedefleyerek CALR ile ilişkili MPN'ler için yeni tedavi stratejilerinin geliştirilmesine yönelik yeni yollar açıyor.

Anahtar kelimeler: Kalretikülin, miyeloproliferatif neoplazmlar, in siliko tahminler, arjinin açısından zengin motifler, hücre dışı sekresyonları teşvik eden sinyaller.



Acknowledgements:

I would like to thank my advisor, Ilyas Chachoua, for his guidance and support throughout my master's program as well as writing our review article for secondary acute myeloid leukemia progression and this thesis. For being jury members for my thesis, I also thank Serkan Belkaya and Ahmet Acar for taking time from their incredibly busy schedules to evaluate this work. Additionally, I would also like to thank my fellow master's student, Beana Puka, for her assistance in running experiments in the lab. Furthermore, I also thank TÜBİTAK for their 3501 grant that has permitted us to start working on our project. Considering that our group was just started in September 2022, credit is also due to the groups of Serkan Belkaya, Pınar Önal, Onur Çizmecioğlu, Özlen Konu, and Serkan Göktuna for their assistance in providing their expertise, lab materials, and time to aid us in getting our lab up and running. Also, we thank Stefan Constantinescu at the Catholic University of Louvain in Belgium for graciously providing the calreticulin plasmids used for mutageneses. Finally, I would also like to extend thanks to my mother, Delayne Auerbach, my wife, Çisil Karagüzel, and her parents Metin and Beyhan Karagüzel, for their support as well throughout the last two years, and I am dedicating this thesis to them.

LIST OF ABBREVIATIONS

AID	<u>Autoimmune disease</u>
alloSCT	<u>Allogeneic stem cell transplantation</u>
APS	<u>Ammonium persulfate</u>
CALR	<u>Calreticulin</u>
CHIP	<u>Clonal hematopoiesis of indeterminate potential</u>
CNN	<u>Convolutional neural network</u>
DAMP	<u>Damage-associated molecular pattern</u>
DOC	<u>Docking site</u>
DTT	<u>Dithiothreitol</u>
DVT	<u>Deep vein thrombosis</u>
EDTA	<u>Ethylenediaminetetraacetic acid</u>
EpoR	<u>Erythropoietin receptor</u>
ER	<u>Endoplasmic reticulum</u>
ET	<u>Essential thrombocytopenia</u>
EZH2	<u>Enhancer of zeste homolog 2</u>
FUR	<u>Furin</u>
HBS	<u>Hanks' Balanced Salt (Solution)</u>
HRP	<u>Horseradish peroxidase</u>
HSC	<u>Hematopoietic stem cell</u>
IDH	<u>Isocitrate dehydrogenase</u>
IL-2	<u>Interleukin-2</u>
IRES	<u>Internal ribosomal entry site</u>
JAK2	<u>Janus kinase 2</u>
JAK-STAT	<u>Janus kinase 2—signal transduction and activation of transcription</u>
KEX	<u>Kexin</u>
LC3	<u>Microtubule-associated protein 1A/1B light chain 3</u>
LFS	<u>Leukemia free survival</u>
LOH	<u>Loss of heterozygosity</u>
MAFFT	<u>Multiple alignments using Fast Fourier transform</u>
MPL/TpoR	<u>Thrombopoietin receptor</u>
MPN	<u>Myeloproliferative neoplasm</u>
MPN-AP	<u>Myeloproliferative neoplasm acute phase</u>
MPN-BP	<u>Myeloproliferative neoplasm blast phase</u>
MPN-CP	<u>Myeloproliferative neoplasm chronic phase</u>
MPO	<u>Myeloperoxidase</u>
NLS	<u>Nuclear localization signal</u>
NP-40	<u>Nonidet P-40</u>
NRD	<u>N-arginine dibasic convertase</u>

<u>OS</u>	<u>Overall survival</u>
<u>OLS</u>	<u>Ordinary least-squares</u>
<u>PAD</u>	<u>Peripheral artery disease</u>
<u>PBS</u>	<u>Phosphate buffer saline</u>
<u>PCSK7</u>	<u>Proprotein convertase 7</u>
<u>PERK</u>	<u>Pancreatic ER kinase</u>
<u>pLDDT</u>	<u>Predicted local distance difference test</u>
<u>PMF/SMF</u>	<u>Primary / secondary myelofibrosis</u>
<u>PRR</u>	<u>Pattern recognition receptor</u>
<u>PSSM</u>	<u>Position-specific scoring matrix</u>
<u>PTM</u>	<u>Post-translational modification</u>
<u>PV</u>	<u>Polycythemia vera</u>
<u>RBC</u>	<u>Red blood cell</u>
<u>ROS</u>	<u>Reactive oxygen species</u>
<u>RMSD</u>	<u>Root mean square deviation</u>
<u>RT</u>	<u>Room temperature</u>
<u>RUNX1</u>	<u>Runt-related transcription factor 1</u>
<u>SDS</u>	<u>Sodium dodecasulfate</u>
<u>SHP</u>	<u>SH2-binding phosphatase family</u>
<u>SOCE</u>	<u>Store-operated calcium (2+) entry</u>
<u>SOCS</u>	<u>Suppressors of cytokine signaling</u>
<u>STIM1</u>	<u>Stromal interaction molecule 1</u>
<u>TAE</u>	<u>Tris base-acetic acid-EDTA</u>
<u>TBS</u>	<u>Tris-buffered saline</u>
<u>TBST</u>	<u>TBS with Tween 20</u>
<u>TET2</u>	<u>Tet methylcytosine dioxygenase 2</u>
<u>TfR/sTfR</u>	<u>Transferrin receptor / soluble transferrin receptor</u>
<u>TNF</u>	<u>Tumor necrosis factor</u>
<u>TP53</u>	<u>Tumor protein p53</u>
<u>TRG</u>	<u>Targeting motif</u>
<u>UPR</u>	<u>Unfolded protein response</u>
<u>VAF</u>	<u>Variant allele frequency</u>
<u>WBC</u>	<u>White blood cell</u>

LIST OF FIGURES

Figure 1. Overview of myeloid cell differentiation inside the bone marrow.....	2
Figure 2. Flowchart showing relationship between different MPN subtypes and sAML.....	4
Figure 3. Role of JAK-STAT pathway in platelet production.....	15
Figure 4. Functions of each domain in calreticulin.....	21
Figure 5. Diagram showing major functions of calreticulin inside the cell.....	23
Figure 6. Calreticulin del52 contributing to MPN development via XBP1.....	31
Figure 7. DeepLoc sorting signals for wild-type and mutant calreticulin.....	56
Figure 8. DeepLoc and MULocDeep predictions for secretion changes for each mutagenesis.....	64
Figure 9. DeepLoc and MULocDeep predictions for cumulative mutations.....	71
Figure 10. MULocDeep predictions for different types of gradients used for K/R tally.....	78
Figure 11. Results for the “extra K/R ” analyses in DeepLoc and MULocDeep.....	88
Figure 12. DeepLoc sorting signal results with modified sequence of wild-type calreticulin.....	98
Figure 13. PyMOL-generated snapshots of calreticulin C-terminus around KKRK motif.....	101
Figure 14. Heatmaps for PyMOL measurements between atoms.....	106
Figure 15. K/R tally versus KDEL tally analyses.....	110
Figure 16. Analyses comparing ins5_alanine with ins5 and del52 mutants.....	115
Figure 17. DeepLoc predictions for comparison of arginines and lysines.....	118
Figure 18. MuSiteDeep post-translational modification predictions for del52 calreticulin.....	121
Figure 19. DeepPhos post-translational modification predictions.....	127
Figure 20. DeepLoc and MULocDeep predictions for select del52 mutations.....	133

LIST OF TABLES

Table 1. WHO characteristics of MPN subtypes.....	6
Table 2. Full sequences of wild-type and mutant calreticulin.....	27
Table 3. Consensus sequences for wild-type and mutant calreticulin.....	35
Table 4. ELM results for wild-type and del 52 calreticulin.....	60
Table 5. Mutagenesis list for both wild-type and del52 calreticulin.....	63
Table 6. Different mutagenesis combinations in wild-type and del52 calreticulin.....	70
Table 7. Example of gradient used for <i>in silico</i> analyses.....	77
Table 8. List of sequences used for “extra K/R” analyses.....	86
Table 9. ELM search results for calreticulin where KKRK motif is mutated to alanines.....	97
Table 10. Statistical correlations for data in Figure 21.....	130
Table 11. Statistical correlations for data in Figure 22.....	131

Table of Contents

Abstract / Özet.....	iii
Acknowledgements.....	vi
List of Abbreviations.....	viii
Lists of Tables and Figures.....	x
1) Introduction.....	1
1.1) Hematopoiesis Overview.....	1
1.2) Myeloproliferative Neoplasm (MPN) Overview.....	3
1.3) JAK-STAT pathway.....	10
1.3.1) JAK-STAT pathway in normal hematopoiesis.....	12
1.3.2) JAK-STAT pathway in MPNs.....	14
1.4) Janus Kinase 2 (JAK2).....	16
1.5) MPL (Thrombopoietin Receptor).....	18
1.6) Calreticulin (CALR).....	20
1.6.1) CALR Structure and Function.....	20
1.6.2) CALR mutants and MPN Disease Progression.....	26
1.6.3) Mutant CALR and the Unfolded Protein Response.....	30
1.7) Aims and Objectives.....	33
2) Materials and Methods.....	38
2.1) Materials.....	38
2.1.1) Reagents for competent <i>E. coli</i> growth and storage.....	38
2.1.2) HEK293 cell media reagents.....	38

2.1.3) List of mutagenesis primers.....	39
2.1.4) List of Western blot reagents (excluding buffers).....	40
2.1.5) List of Western blot buffers with individual ingredients.....	41
2.1.6) Mutagenesis, plasmid miniprep kits, DNA measurements, and agarose gels.....	42
2.1.7) List of plasmids used for mutagenesis.....	43
2.2) Methods.....	43
2.2.1) DeepLoc sorting signal prediction.....	43
2.2.2) Eukaryotic Linear Motif search engine tool.....	44
2.2.3) Localization predictions with DeepLoc and MULocDeep.....	44
2.2.4) Preparation of competent cells for bacterial transformation.....	45
2.2.5) Site-directed mutagenesis of CALR del52 truncations.....	46
2.2.6) Miniprep procedures for mutagenesis colonies.....	47
2.2.7) Sanger sequencing result analysis with Benchling.....	49
2.2.8) Calcium phosphate transfection of HEK293 cells.....	50
2.2.9) Western blot protocol.....	51
2.2.10) Obtaining PDB files for CALR and visualization with PyMOL.....	53
2.2.11) Predicting PTM on select residues on CALR del52.....	54
3) Results.....	55
3.1) DeepLoc sorting signal results.....	55
3.2) ELM search results for CALR del52.....	59
3.3) Individual motif mutageneses for CALR WT and del52 in DeepLoc / MULocDeep....	62
3.4) Combinations of del52 and WT mutageneses in DeepLoc and MULocDeep.....	68

3.5) K/R tally gradient approach to predicting levels of CALR del52 secretion.....	76
3.6) Addition of K/R residues to CALR del52 towards either terminus.....	85
3.7) Examination of KKRRK motif in the CALR WT sequence.....	96
3.8) Visualization of CALR PDB files with PyMOL.....	100
3.9) Distance heatmaps for residues flanking CALR WT KKRRK.....	105
3.10) Comparing effects of K/R tally versus KDEL tally on predicted del52 secretion.....	109
3.11) Comparison of CALR ins5 and CALR ins5_alanine.....	114
3.12) Comparison of lysines and arginines' ability to affect CALR del52 secretion.....	117
3.13) Investigation of possible correlation between PTM and K/R tally.....	120
3.14) Examination of loss of potential PTM on CALR del52 secretion.....	132
4) Discussion.....	137
5) Future Directions.....	139
6) References.....	140



CHAPTER 1

INTRODUCTION

1.1) Hematopoiesis Overview

Hematopoiesis is the development of blood cellular components that is initiated during embryogenesis and continues until adulthood, and is composed of two waves: the primitive wave and the definitive wave. The former involves the development of, among other cell types, erythroid progenitors and macrophages in the embryo, while the latter is characterized by development of myeloid progenitors in addition to erythroid progenitors. In the case of humans, these progenitors are exclusively generated in the bone marrow from about 10.5 weeks after gestation, after some initial activity in the fetal yolk sac and the liver [1]. There are two main pathways of hematopoiesis that counterbalance each other in typical homeostasis: self-renewal of hematopoietic stem cells (HSCs) and differentiation into a varied repertoire including red blood cells (RBC), white blood cells (WBC), and platelets. This balance is maintained with the assistance of cytokines like thrombopoietin [2], but is not always successful. A disproportionately frequent phenomenon in elderly patients is clonal hematopoiesis, or the expansion and eventual dominance of a single HSC through acquired mutations. Typically, by the age of 70, humans have obtained between 350,000 to 1,400,000 mutations across their HSC populations, and any of these is potentially significant to start the mutational cascade that may lead to this dominance.

10% of patients with known cytopenia or blood disorders over the age of 60 have clonal hematopoiesis of indeterminate potential (CHIP), where HSCs have a variant allele frequency (VAF), or the prevalence of a variant allele among all genomic reads, of 0.02 [3]. The risk of developing any type of hematological malignancy increases by 0.5% to 1% every year [4]. This elevated frequency is in some cases sufficient to result in development of several types of myeloid neoplasms, including but not limited to myeloproliferative neoplasms.

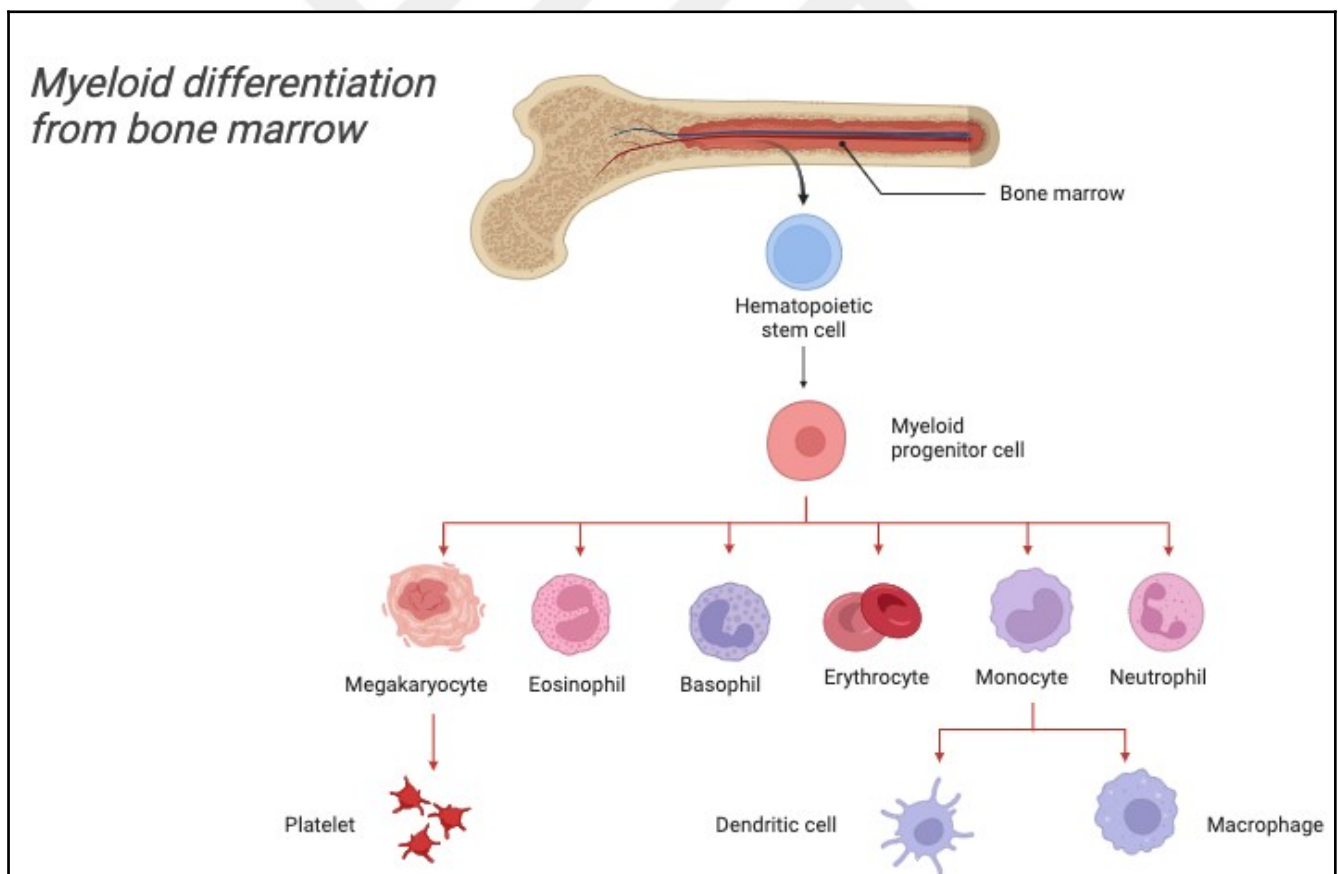


Figure 1. Overview of the development of different myeloid cell types from HSCs and myeloid progenitors (red arrows indicate differentiation into specific cell lines). Image generated by BioRender.

1.2) Myeloproliferative Neoplasms (MPNs) Overview

Myeloproliferative neoplasms (MPNs) are heterogeneous hematopoietic clonal disorders characterized by abnormal hematopoietic stem cell (HSC) transforming myeloid progenitor cells and enhancing their ability to proliferate and divide into multiple types of myeloid cells. There are two types of MPNs, BCR-ABL (or Philadelphia chromosome) positive and negative, with the former referring to chronic myeloid leukemia, which is also included in the MPN category, and the latter referring to MPNs with the potential to progress to secondary acute myeloid leukemia (sAML) [1].

There are three main subtypes of MPNs: polycythemia vera (PV), essential thrombocythemia (ET), and myelofibrosis (MF). They differ in terms of the cell types amplified: PV patients have more RBC, whereas ET and PMF patients have more platelets. According to WHO guidelines (see Table 1 for further details), the vast majority of PV patients are characterized by homozygous *JAK2* mutations in exon 12 (V617F mutation, 1849G>T), while ET and PMF patients are also characterized by mutations in either calreticulin (*CALR*) or *MPL* (myeloproliferative leukemia protein, named as such because of its association with the disease) in addition to *JAK2* mutations [3]. Despite all three mutations having proven contributions to MPN genesis and possible subsequent transformation to sAML, triple negative MPN patients have significantly worse overall prognosis compared to patients who had mutations in any of the three genes [4, 5]. These three subtypes are not exclusive to each other within the same patient: it is possible for a patient with either PV or ET to develop a form of myelofibrosis, referred to as *secondary* myelofibrosis (SMF) [6]. For MF patients (either PMF or SMF), the OS rate, and particularly leukemia-free survival (LFS), is worse compared to either PV or ET: 60% for patients with less than 5% blasts and 30% for patients with $\geq 5\%$ blasts. In addition to blast

count, both RBC, especially post-transfusion, and WBC counts were positively correlated with LT [37].

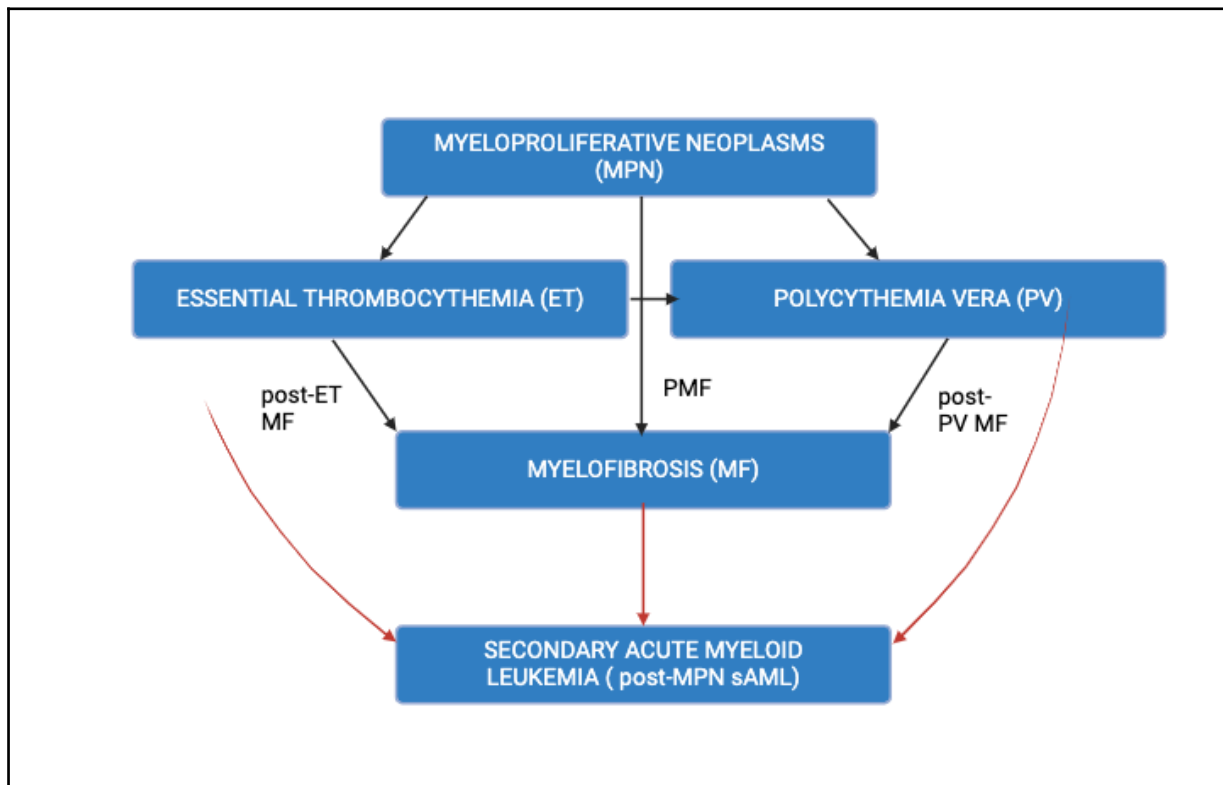


Figure 2. Flowchart showing different subtypes of MPNs, their relationship to each other, and progression to sAML (shown with red arrows). Image generated by Biorender.

MPN aggressiveness can be divided into three phases based on their blast percentage in the bone marrow or peripheral blood: 0-9% is the chronic phase (CP), 10-19% is the accelerated phase (AP), and $\geq 20\%$ is the blast phase (BP) according to 2016 WHO guidelines. Once a MPN patient has reached the blast phase, they are diagnosed with sAML. This process takes years and most MPN-CP or

MPN-AP patients do not ever surpass the threshold. Those that do reach BP first go through MPN-CP and then MPN-AP.

The first way to distinguish between MPN subtypes according to WHO criteria is the relative amount of different cell types and characteristics: ET patients have a relatively high platelet count, PV patients have increased red blood cell mass, and MF (including PMF, post-ET SMF, and post-PV SMF) patients have noticeable splenomegaly (spleen enlargement) and enhanced fibrosis inside the bone marrow. MPNs are also distinguished from sAML by the myeloid blast percentage in the bone marrow or peripheral blood: if a MPN patient surpasses the 20% threshold, they are diagnosed with sAML [33]. Most MPN cases are sporadic and seen in older adults, especially those older than 60, although there are rare germ-line mutations passed down from patients as well. One major factor is CHIP, or when a mutation associated with myeloid malignancies has an allele frequency of over 2%. Not everyone with CHIP will develop these malignancies, but they are at a significantly higher risk of doing so. This is because there are heightened levels of inflammation, and these mutant cells have a selective advantage that allows them to gain clonal dominance over time. Mutations in genes that affect epigenetic modifications like *TET2* or *DMNT3A* also facilitate these clones' growth over time. Later on, any of the three key driver mutations emerging in the cells, particularly *JAK2* V617F, also confer an advantage in the inflammatory environment [17]. Inflammatory cytokines also facilitate the transition from either PV or ET to SMF, especially interleukin-2 (IL-2) and IL-6, although they usually work in tandem with either TNF or *JAK2* V617F mutation, so they are not solely responsible. As people age, there is also an increase in general inflammation stemming from heightened TNF activity, which is another reason MPN patients tend to be older rather than younger [12].

Table 1. Display of different criteria and driver mutations associated with each MPN subtype as established by the 2016 World Health Organization standards [8]. Major criteria are the most likely features to be observed with each MPN subtype, while minor criteria are less commonly observed. Genes with driver mutations that contribute to MPN development are also included.

MPN Subtype	PV	ET	MF
Major Criteria	Hemoglobin ≥ 16 g/dL, hematocrit $> 48\%$, increased RBC mass, prominent granulocyte / megakaryocyte growth	Platelet count $\geq 450 \times 10^9/L$, megakaryocyte proliferation	Megakaryocyte proliferation, absence of reactive myelofibrosis
Driver Mutations	JAK2 V617F, exon 12 mutations	JAK2, CALR, MPL	JAK2, CALR, MPL
Minor Criteria	Subnormal serum erythropoietin level	Absence of evidence for reactive thrombocytosis	Leukocytosis $\geq 11 \times 10^9/L$, palpable splenomegaly

Inflammation also plays a key role in MPN genesis, as seen when allogeneic stem cell transplantation (alloSCT) removes MPN cells and eventually replaces them with healthy hematopoietic cells, removing the inflammatory bone marrow environment and successfully treating MPNs. In particular, experiments in mice transfected with *MPL* mutant (one of the key driver mutations of MPNs), cells have higher levels of inflammatory cytokines such as but not limited to IL-6, IL-10, and TNF α (tumor necrosis factor alpha). Additionally, the overproduction of lymphocytes and platelets from abnormal myeloid progenitor cells further leads to further inflammatory cytokine recruitment and sustenance of a bone marrow microenvironment that has the potential to progress to cancer, although most MPN patients do not progress to that stage [7]. People with a history of autoimmune diseases are also at a higher risk of developing MPNs because of the elevated level of pro-inflammatory cytokine activity prior to any mutations in any of the genes listed above. TNF also helps establish clonal dominance of aberrant myeloid progenitor cells that give rise to MPNs by suppressing benign myeloid

progenitor proliferation, allowing them to be gradually replaced by MPNs. These MPNs further stimulate inflammatory cytokines, and thus TNF activity, resulting in a positive feedback loop that accelerates their clonal dominance [7].

Another related factor that increases the likelihood of developing MPNs is autoimmune diseases (AID): patients who have this type of disease already have heightened inflammation, increasing the possibility of acquiring CHIP. Although only approximately 2% of MPN patients had prior AID, they have a 20% higher likelihood than those without AID of developing MPNs. Without any detectable antibodies, the risk is 60% higher. This varies by disease, ranging from an odds ratio of 2.9 for thrombocytopenic purpura, or autoimmune targeting of platelets, to 7.8 for aplastic anemia [34]. The most common AIDs associated with MPNs are arthritis, systemic vasculitis, and dermatosis. In terms of mutation acquisition, there is generally no significant difference between MPN patients with or without an antecedent autoimmune disease, with the exception of *TET2* (tet methylcytosine dioxygenase 2). MPN patients with antecedent AID were far more likely to acquire *TET2* mutations. Even though they are more likely to develop MPNs, patients with AID had similar OS rates and SMF (post-EV or post-PV) survival rates. AIDs only accelerate the development of MPNs, but do not make a difference afterwards [35].

Symptoms associated with MPNs are fatigue, splenomegaly, dizziness, night sweats, and elevated myeloid cell levels [9]. Co-morbidities common in MPN patients are cardiovascular and cerebrovascular disorders, which cause almost half of all fatalities in MPN patients rather than the MPNs themselves. These disorders range from relatively benign deep vein thrombosis (DVT) to potentially lethal myocardial infarctions. Compared to the general population, MPN patients are less likely to suffer from strokes or myocardial infarctions, but more likely to have either peripheral artery

disease (PAD) in the lower extremities or DVT. The overall five-year hazard ratio for MPN patients compared to the general population was 1.3 to 3.7. Heavy smokers' blood samples can be similar to those of MPN patients in terms of composition to the point where precautions have to be taken to avoid mixing the two types of blood samples [9].

The mutations in the driver genes listed above are somatic and thus not hereditary, although people with two or more family members diagnosed with MPNs are at a higher risk than the general population of developing it as well. The existence of other germ-line mutations in these key driver genes that have a similar effect to those somatic mutations already mentioned have been hinted at, but they constitute a minority of MPN patients. It is also difficult to ascertain whether these patients actually have some form of familial MPN or instead have a similar disorder like familial erythrocytosis, which is also characterized by elevated red blood cell counts. The main difference between the two is the set of germ-line mutations in familial erythrocytosis, found in *JAK2* and *MPL* but not *CALR*. At most, 7% of MPN patients have a family history, while the remainder acquire it through the aforementioned somatic mutations [11, 12]. The main focus thus will be on the somatic mutations, which will be elaborated on in further sections.

The incidence rates for each subtype somewhat vary, but they all are still relatively rare: 0.84, 1.03, and 0.47 per 100,000 for PV, ET, and PMF, respectively. SMF is slightly more common (at least in the United States) than PMF with an incidence rate of 0.5 per 100,000 for post-PV SMF and 0.8 per 100,000 for post-ET SMF, respectively [14]. There is also the issue of progression to secondary AML. Based on a study by Hong et al on MPN patients in South Korea, 3.6% of ET patients, 1.7% of PV patients, and 21.4% of PMF patients were observed to progress to sAML over the course of eight years. Intriguingly, SMF patients were significantly less likely than PMF patients to progress to sAML: only

6.8% of patients with SMF progressed to sAML over a four-year period. Although no long-term data was collected for SMF patients in Korea, European studies have estimated that 10-20% of SMF patients progressed to sAML over the course of ten years, which is still lower than that of PMF patients. Overall survival (OS) rates ten years after diagnosis for MPN patients vary: 60-83% for ET patients, 56-83% for PV patients, and 21-46% for PMF patients [15, 16]. Given that the median age of MPN patients ranges from 65-70 depending on the subtype [17], the overall picture is that MPN patients live out the duration of their lives managing the disease while always monitoring for the possible progression to sAML.

Although most MPN patients do not progress to sAML, this section will address the minority that does. In addition to the aforementioned three driver mutations, mutations that facilitate leukemogenesis are most commonly found in the genes *TP53* (tumor protein p53), *TET2*, *RUNX1* (Runt-related transcription factor 1), *ASXL1* (additional sex combs 1), and *EZH2* (enhancer of zeste homolog 2). Some of these mutations confer either a faster or slower timeframe of leukemic evolution: *TP53* mutations were more likely to be observed in patients that took at least ten years to progress to sAML, while mutations in either *IDH1/2* (isocitrate dehydrogenase 1 or 2), or *RUNX1* were more common in ET/PV patients who progressed to sAML roughly three years after diagnosis [20]. *IDH1/2* mutations are typically loss-of-function in MPNs, and their loss means that HSCs cannot differentiate into myeloid cells and continue proliferating instead, eventually leading to sAML development. This process is accelerated further when the *JAK2* V617F is present, which occurs in most MPNs [21]. *RUNX1* (also known as *AML1* due to its association with the disease) has a point mutation, D171N, that promotes HSC self-renewal and proliferation rather than differentiation. This mutation is also amplified

by the pro-proliferation JAK-STAT pathway, which is further enhanced by the *JAK2* V617F mutation [22].

1.3) The JAK-STAT Pathway

The Janus kinase/signal transducers and activators of transcription (JAK-STAT) is one of several mammalian signaling pathways responsible for cell proliferation, differentiation and survival. It is used for many types of cells, including but not limited to hematopoietic, adipose, and mammary. When a ligand binds to its receptor on the cell surface, for example erythropoietin, the receptor alters its configuration and is multimerized to recruit JAK kinases. The JAK kinase is then phosphorylated and in turn phosphorylates either the corresponding downstream part of the receptor and a STAT protein by creating a docking site using a phosphorylated tyrosine. Typically, STAT proteins reside idly in the cytoplasm, but relocate to the nucleus thanks to the Ran nuclear import pathway after docking the activated JAK. Similar to cell surface receptors, STATs are also dimerized once phosphorylated to facilitate nuclear localization. Once in the nucleus, these STATs either repress or activate transcription of target genes [25, 26].

Given its central role in cell proliferation, the JAK-STAT pathway has several negative regulators including receptor internalization inside the cell membrane, de-phosphorylation of kinases using tyrosine phosphatases, and direct inhibition of STATs by suppressors of cytokine signaling (SOCS). SOCS recruit the E3 ligase complex to ubiquitinate the bound proteins and mark them for degradation. The SH2-binding phosphatase family (SHP) either targets the SH2 domain in the JAK family structure or an analogous domain in the STAT structure, so they are capable of stopping the pathway at either the cytoplasm or the nucleus just before STAT target gene transcription [27].

Conversely, there is also positive regulation of the JAK-STAT pathway that increases the pro-transcription activity of STAT proteins depending on the needs of the cell. For example, the cytoplasmic Nmi protein is able to bind to facilitate activation of either STAT3 or STAT5 by CBP, or CREB-binding protein, which facilitates nuclear localization [28, 29]. Interferon receptors 1 and 2 also bind to JAK kinases to phosphorylate STAT1, which in turn moves to the gamma interferon activation site (GAS) to stimulate transcription of that interferon as part of the immune response. Although up-regulation of JAK-STAT activity does play a role in tumorigenesis, it also participates in activating immune surveillance that works to prevent tumor development and metastasis [30].

Thanks to the JAK and STAT families having several members each, there are approximately forty different cytokine receptors that are capable of coordinating with them to trigger activation of selected genes. Out of the four main members of the JAK family (JAK1, JAK2, JAK3, TYK2), most cytokine receptors (particularly interleukins) use a combination of the first three. There are some interleukins like IL-12 and IL-23 that employ TYK2, showcasing the versatility of the pathway. The STAT family has seven members, and like the JAK kinases, each of them are employed by different types of tissues. Most cell types regulate transcription using STAT1, while macrophages and B/T cells recruit STAT6 instead. Non-immune cells, on the other hand, like hematopoietic and mammary epithelial cells, are more associated with STAT5, although in the case of the former, STAT3 and STAT1 also play a role in cell proliferation, or if needed, repression of apoptosis. Cytokine receptors involved in blood cell production like erythropoietin and thrombopoietin recruit JAK2 as opposed to JAK1 or JAK3. These receptors also differ from interleukins in that they activate the pathway through homo-dimerization as opposed to coordination of multiple chains. [31–33].

1.3.1) JAK-STAT Pathway in Normal Hematopoiesis

In early hematopoiesis, the JAK-STAT pathway is not only crucial for HSC development and proliferation, but also self-renewal and maintenance. Among the STAT family, STAT3 supports the former while STAT5 is more associated with the latter [34]. In myeloid cells, the JAK-STAT pathway primarily involves STAT5 being activated while STAT3 is sometimes employed, although there is a delicate balance between which STAT is utilized. The interplay between these two STATs is crucial for maintaining normal hematopoiesis: excess STAT5 activity versus STAT3 has been associated with progression to leukemia [35].

Human and murine HSC development through the JAK-STAT pathway is negatively regulated in particular by GATA-2, a downstream transcription factor that maintains HSC in the G₀ phase, also known as quiescence, thus preventing differentiation and proliferation [36]. Quiescence is also facilitated by p53 activity and the tumor suppressor PTEN. On the other hand, self-renewal is suppressed by Lnk, which directly acts on JAK2 and prevents it from binding to STAT5, which is associated with quiescence as well as cell proliferation, depending on its target, and thus is capable of suppressing or enabling HSC differentiation depending on the needs of the cell. Even in the presence of c-MPL (myeloproliferative leukemia protein, named as such because of its association with the disease), STAT5 is still capable of maintaining steady-state hematopoiesis. [37]. The biggest regulator of the JAK-STAT pathway is the nature of the cytokine receptor: without corresponding ligand binding, the receptor does not dimerize, and thus the JAK is not phosphorylated [38].

An example of this is the lack of STAT5 activity in the presence of IL-6 as opposed to thrombopoietin – even when phosphorylated at the correct tyrosine residues, STAT5 remained in the cytoplasm and was functionally inactive. However, when thrombopoietin was present, the

phosphorylated STAT5 successfully moved to the euchromatic regions in the nucleus to target downstream genes related to proliferation. Unlike STAT5, STAT3 is correctly stimulated by IL-6 in conjunction with gp130 – and this results in HSC differentiation into myeloid cells rather than self-renewal [39]. STAT3 activation buoyed by JAK phosphorylation resulting from either IL-6 or IL-11 is vital for preventing proliferation of immature HSCs, but too much often results in the development of myeloid cancers [40]. STAT1 also plays a role in adult hematopoiesis, but is much more specific as a response to stressful events like blood loss or an acute infection while working among quiescent adult HSCs.

JAK proteins excluding TYK2 are associated with different types of hematopoietic cells due to coordinating with different STAT proteins: JAK1 is responsible for lymphoid development, as is JAK3, which also promotes myelopoiesis. JAK2, on the other hand, is exclusively intertwined with myeloid development as seen by its activation by either thrombopoietin or erythropoietin. Mouse knockout studies have demonstrated the crucial role of JAK2 for long-term survival. Embryos that had no functional JAK2 were no longer viable even if other cell type development, like that of lymphoids, was left intact. Other contributions to this lethality were absent cytokines like MAPK and PI3K that also contribute to overall cell survival [41]. Similar to other tyrosine kinases, the JAK family also causes diseases, particularly hematopoietic ones like MPNs, when they are constitutively activated. As opposed to normal hematopoiesis, this is not halted by the absence of the MPL or erythropoietin ligands [35].

1.3.2) JAK-STAT pathway involvement in MPNs

JAK signaling is typically involved in cell proliferation and differentiation, and the immune response. Although it is mostly associated directly with STAT dimerization or polymerization, it is also involved in activating other kinase pathways such as MAPK and PI3K. [29]. Normally, JAK2 is only active when phosphorylated, but the V617F mutation permits it to remain constitutively active and in turn phosphorylate STAT5 and activate it through homodimerization. The dimerized STAT5 then shuttles to the nucleus and acts as a transcription factor to regulate cell proliferation (see Figure 3). [30].

STAT 3/5 are also dimerized as an effect of being phosphorylated by JAK2, and especially the V617F mutant. The main effect of STAT 3/5 dimerization, and then subsequent tetramerization in the case of the latter, is the STAT attachment to importins and relocation to the nucleus where it acts as a transcription factor to upregulate cell proliferation. This means of dimerization/tetramerization is one of the mechanisms that cause leukemogenesis from MPNs [33]. The importance of the JAK-STAT pathway has been emphasized even further by monitoring MPN patients when it is inactive by treating them with the JAK2 inhibitor ruxolitinib. In MF patients in particular, the inhibitor has been shown to significantly improve patients' outcomes [34].

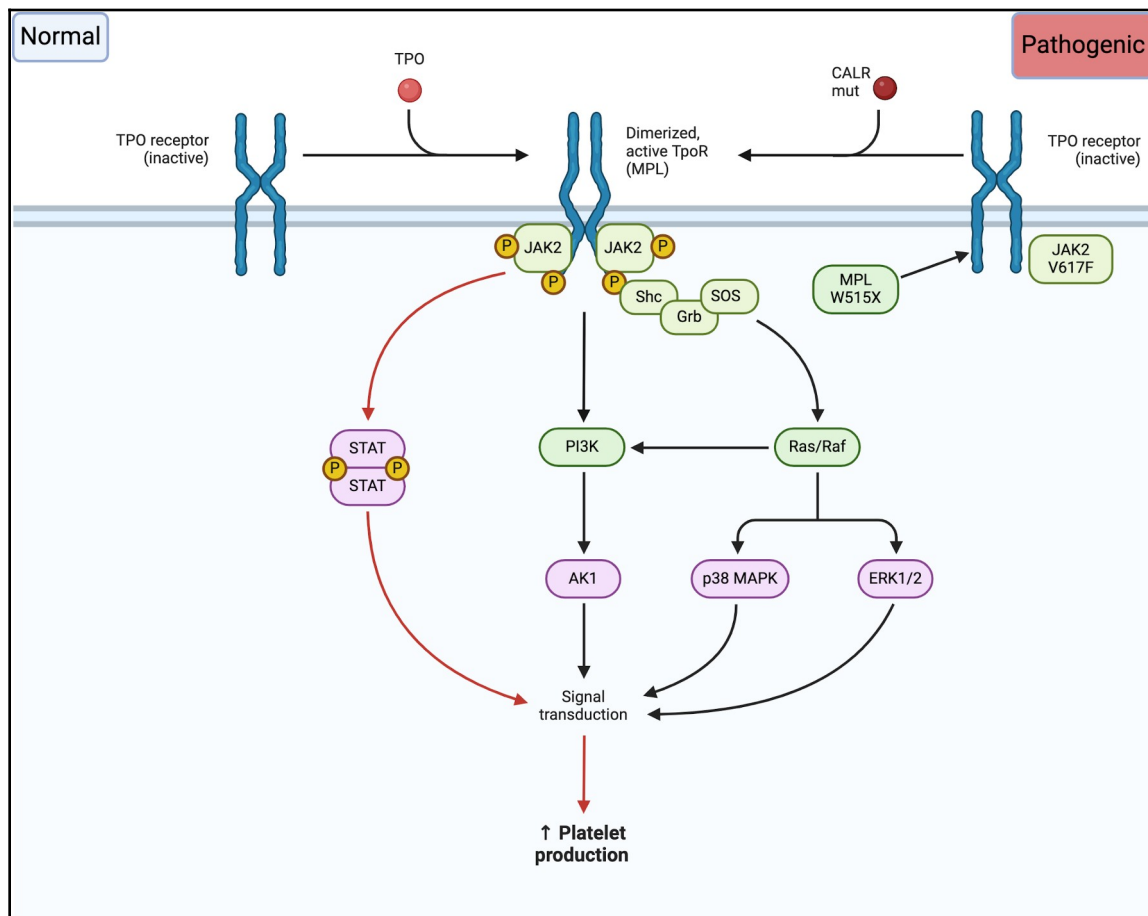


Figure 3. Role of JAK-STAT pathway (STAT dimerization highlighted in red) initiated by MPL binding to its receptor in platelet activation [35]. Image created using templates in Biorender. Normal pathway activation shown on top left, pathway activation in MPN patients shown on top right. On the top right, the approximate location of the W515X mutation on MPL shown with an arrow.

Another reason for the constitutive activity of the JAK-STAT pathway that stems from the *JAK2* V617F mutant is the inability of the wild-type JAK2 negative regulator SOCS3 (suppressor of cytokine signaling 3) to block V617F phosphorylation of STATs, allowing the process to continue unabated. Finally, phosphorylated serines on STATs prevent another type of post-translational modification (PTM), dimethylation by EZH2 on residue K46 which hinders gene transcription. When this is prevented, IL-6 up-regulates STAT3 activity [35]. IL-6 activity is up-regulated through the

CALR del52 mutant, named after a 52-bp frameshift deletion in the C-terminus, and *JAK2* V617F, and when its receptor is lost, MPN formation is halted even in the presence of the aforementioned *CALR* and *JAK2* mutations [36]. The amplifying effect of *IL-6* in the absence of *MPL* mutations, but not *CALR* or *JAK2* also helps explain why not all three mutations have to occur simultaneously for MPN genesis and to a lesser extent, progression to sAML.

There are other pathways affected by amplified CALR del52 activity, and subsequently MPL and JAK2 mutant activity, including but not limited to MAPK and PI3K, but they are not the focus of this thesis. Alternative pathways such as MAPK and PI3K also contribute to increased cell proliferation in MPNs and sAML that do not necessarily involve the exact same compounds, but they all have more or less the same effect on MPNs and the overall prognosis of the patient in question.

1.4) Mutations in Janus Kinase 2 (JAK2)

The most frequent mutation in *JAK2* is V617F at a residue located in its homology domain 2 (JH2), which regulates its tyrosine kinase domain (JH1) that is capable of auto-phosphorylation. This mutation is simultaneously a loss-of-function as well as gain-of-function mutation, or more aptly, a switch-of-function in that it enables rather than prevents oncogenic behavior. As V617 is mutated to a phenylalanine, the spatial structure of the JH1 domain is altered such that two residues gain the ability to phosphorylate themselves and this arrangement becomes energetically favorable [49]. Once auto-phosphorylated, JAK2 is able to activate downstream STAT signaling independent of antecedent cytokines. This mutation is seen in the majority of MPN patients, including all three main types of MPNs. There are also other mutations found in *JAK2* exon 12 that co-occur with V617F [50]. The JH2 domain is also considered to be a pseudo-kinase domain due to its lack of active catalytic sites. This is

a result of phenylalanines blocking the catalytic regions from being activated, but this inhibitory mechanism is halted when *JAK2* acquires the V617F mutation [51]. A section of the JH2 domain, known as alpha-C, loses its ability to alternate between an active conformation and an inactive one, due to the active conformation gaining energetic favorability as a result of rearrangement of the JH2 from the newly acquired, bulky phenylalanine residue, resulting in higher kinase activity [52]. *JAK2* mutations are mutually exclusive with either *CALR* or *MPL* mutations, and do have distinct clinical characteristics: they are more common in older patients and females for both ET and PMF compared to either *CALR* or *MPL* mutations, and do have a sizable impact on the overall survival time of patients.

In addition, the type of MPN accompanying the mutation dramatically affects the long-term patient prognosis. Compared to *CALR* mutant PMF patients, *JAK2* mutant patients had a much shorter survival time (4.8 versus 15.5 years), but for ET patients the median OS times were statistically identical: 19.7 years for *CALR* mutant patients and 19 years for *JAK2* mutant patients [53]. Other effects of the *JAK2* V617F mutation seen in the clinic are increased production of red blood cells, granulocytes, and platelets [54]. In addition to accelerating cell proliferation, *JAK2* V617F also maintains a state of chronic inflammation by up-regulating cytokine activity and production of reactive oxygen species (ROS). Since it is the most common mutation in MPN patients and the first to arise, it is highly likely that the mutation occurs in childhood and thus the mutation confers a selective advantage to its clones allowing them to establish dominance over time [12].

JAK2 V617F also has a modest impact on initiating inflammatory cytokines: there is a positive correlation between its activity and that of IL-6 and TNF α , but when it is knocked down the latter two are still active, so it is not strictly necessary for their activation [55]. As mentioned before, mutant *JAK2* up-regulates STAT multimerization and activation, but it also affects other pathways that have a

similar effect on cell proliferation. Other genes amplified by V617F include but are not limited to the oncogene *C-myc*, DNA repair protein Rad51, and the transcription factor *ID1* that promotes myeloid differentiation [56]. Similar to either the del52 or ins5 mutations in *CALR*, *JAK2* V617F is always homozygous, but loss of heterozygosity (LOH) does occur with loss of chromosome 9p during aberrant mitotic recombination. Another difference between the two is that homozygous *JAK2* V617F can bind to and activate both MPL and EpoR whereas heterozygous *JAK2* V617F only binds to and activates the former [57]. The chronological order of mutation acquisition also affects which MPN subtype the patient has: if V617F precedes mutations in either *TET2* or DNA methyltransferase *DNMT3A*, the patient acquires PV, otherwise the patient acquires ET [22, 57, 58].

1.5) Mutations in Thrombopoietin Receptor (MPL)

The main mutation in *MPL* that occurs in MPNs is W515L (c.1544G>T) with a less common but a functionally equivalent variant W515K (c.1543T>A or c.1544G>A). It tends to not occur concurrently with the *JAK2* V617F mutation, but there are rare cases where a patient has both: the latter mutation tends to appear later as a “secondary” mutation [59]. There are other mutations that are far rarer but happen at the same residue or very close to it: W515A, W515S, and S505N. These mutations are gain-of-function and promote cytokine-independent cell growth specifically by triggering the transition from G1 to S in the cell cycle [60]. Tryptophan 515 is located in the juxtamembrane intracellular domain of MPL and when mutated to a lysine or a leucine, it results in a constitutively active dimer. This switch from the loss of the tryptophan makes the dimer formation more energetically favorable compared to that of a monomer, facilitating the formation of the dimer and preventing a reversal to a monomer form. S505N (c.1514G>A) has a similar effect, but is not as frequently observed

in patients as W515 mutations. An effect of MPL dimerization is the activation of STATs, which then relocate to the nucleus and act as a transcription factor [61].

Without the W515 mutation, S505C (c.1514G>A) is not enough to trigger dimerization, so it still acts as a passenger mutation rather than a driver[62]. There are other mutations that affect dimerization as well: L498W, H499C, and H499Y all of which can trigger dimerization independent of W515. The first two have a synergistic effect on TpoR activation based on luciferase levels, but unlike the W515 residue, are not located in the cytosolic region of the receptor but rather the transmembrane region instead. However, all of this dimerization had been stopped by mutating a tryptophan at position 491 to an alanine: the tryptophan rotates from near the cell surface to the transmembrane region allowing monomers to join, but the alanine could not do any of that, preventing dimerization. The exact mechanism involves breaking apart the helices and providing access to S505 to get phosphorylated, starting the cascade that leads to STAT activation and gene transcription. However, if the helices are never broken, none of this happens and the MPL monomers remain inactive [63]. A 2020 deep mutational scanning study of MPN patients by Bridgford et al. found that nearby hydrophobic transmembrane residues being mutated also led to constitutive activation: V501A, L500V, and V507I. Most mutations in this transmembrane exon 10 region were deleterious in terms of MPL activation, but the main exception to this were a select few residues being non-polar granting an edge in terms of proliferative ability, as measured by heightened ERK activity and overall cell growth. Given that these newly identified mutations have been labeled as driver mutations alongside W515K, it is possible that they also disrupt the helicity of the transmembrane region, permitting access to S505 [64].

1.6) Mutations in Calreticulin (CALR)

1.6.1) Calreticulin Structure and Functions

Originally characterized as a calcium-binding protein in muscular tissues, CALR was later found to be ubiquitous through various cell types and species [71]. In humans, it has an approximate weight of 47 kDa, is 411 amino acids long, and is found on chromosome 19 with nine exons. It is composed of three domains, with each covering approximately a third of the sequence length: the N-domain that includes a 17-residue signal sequence directing it to the endoplasmic reticulum (ER), the proline-rich P-domain hosting the calcium binding site, and the C-domain containing the KDEL anchor sequence (see Figure 4) [72]. Apart from the signal sequence, the N-domain interacts with alpha-integrins and the DNA-binding region of the steroid hormone. It also has a globular structure that assists with folding immature proteins in the ER. Most of the protein folding occurs in the P-domain thanks to the repetitive serine and threonine-rich motifs that it contains. The C-domain also plays a role in protein folding thanks to the prevalence of acidic residues, which also help with calcium binding. However, there is a lower affinity for calcium compared to the P-domain, but the capacity is higher because of the longer acidic motifs [73]. CALR has a flexible structure in solution that permits it to bind to various proteins inside the ER as their chaperone to facilitate proper folding and maturation. The N-domain folds into a globular structure when bound to the target protein in order to shield it from proteolysis [74].

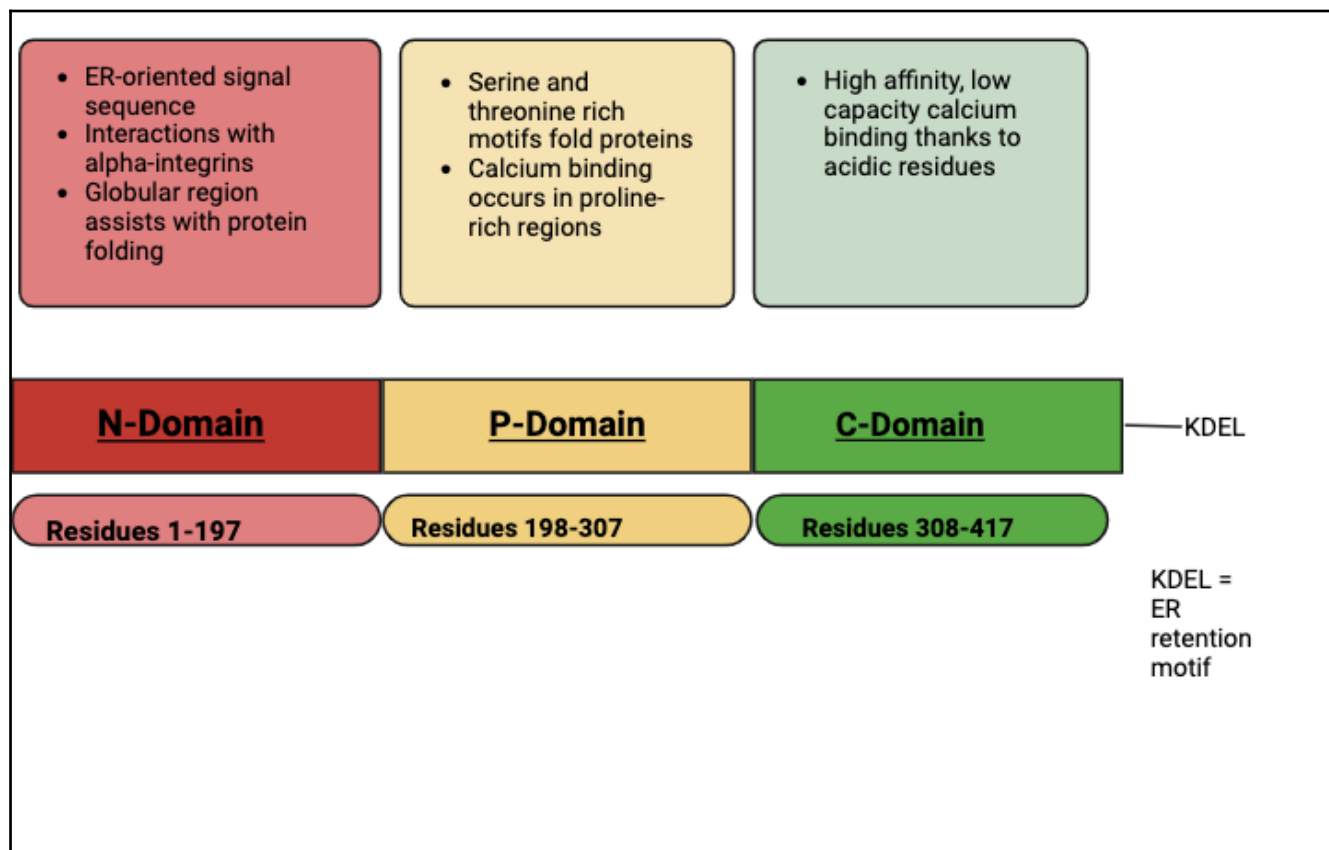


Figure 4. List of functions for each CALR domain plus KDEL ER retention motif and their location along the protein sequence [73]. Image generated by Biorender.

The ER is where protein folding and maturation take place, specifically in the lumen, and CALR fulfills this role in tandem with calnexin and ERp57. All proteins that pass through the ER when they are secreted bind to this complex. CALR specifically targets the monoglucosylated carbohydrates on these newly synthesized proteins, and since this type of interaction is dependent on calcium levels in the ER lumen, the calcium-binding regions of CALR also play a role to ensure proper glycoprotein folding. However, calcium is not always available in sufficient amounts, and there are also proteins that are not glycosylated while passing through the ER, so zinc is used in lieu of calcium to facilitate

folding of non-glycosylated proteins. The P-domain is analogous to an extended arm that brings in the ERp57 into the protein-folding complex in such a way that there is a sizable cavity that accommodates the incoming immature protein [75].

In addition to, and often in conjunction with, its role as a protein-folding chaperone, CALR also plays a role in calcium storage in the ER lumen. The C-domain has a high binding affinity for calcium and has the capacity to sequester up to half of all free calcium inside the ER lumen. Balanced calcium concentrations result in maintained traffic from the ER to the Golgi membrane and proteins moving across the nuclear pore. Unlike the aforementioned chaperone complex, CALR does not work in tandem with other proteins, as seen in knockout studies where CALR-deficient cells had substantially lower calcium levels in the ER but similar levels of folded mature proteins [76]. As a consequence of modulating calcium concentrations, CALR also affects apoptosis and cell adhesion (see Figure 5) [75]. Using the captured calcium, CALR binds to alpha-integrins at their cytoplasmic domain, which then allows the cell in question to adhere to the extracellular matrix [77]. Although CALR is capable of relocating to other organelles, it performs this function mostly in the ER because of its KDEL retention motif and utilizes signaling pathways oriented towards the nucleus and cell surface. Additionally, CALR promotes fibronectin mRNA and protein expression and focal contact assembly through the NFAT/calcineurin and IP3/PTEN pathways [78].

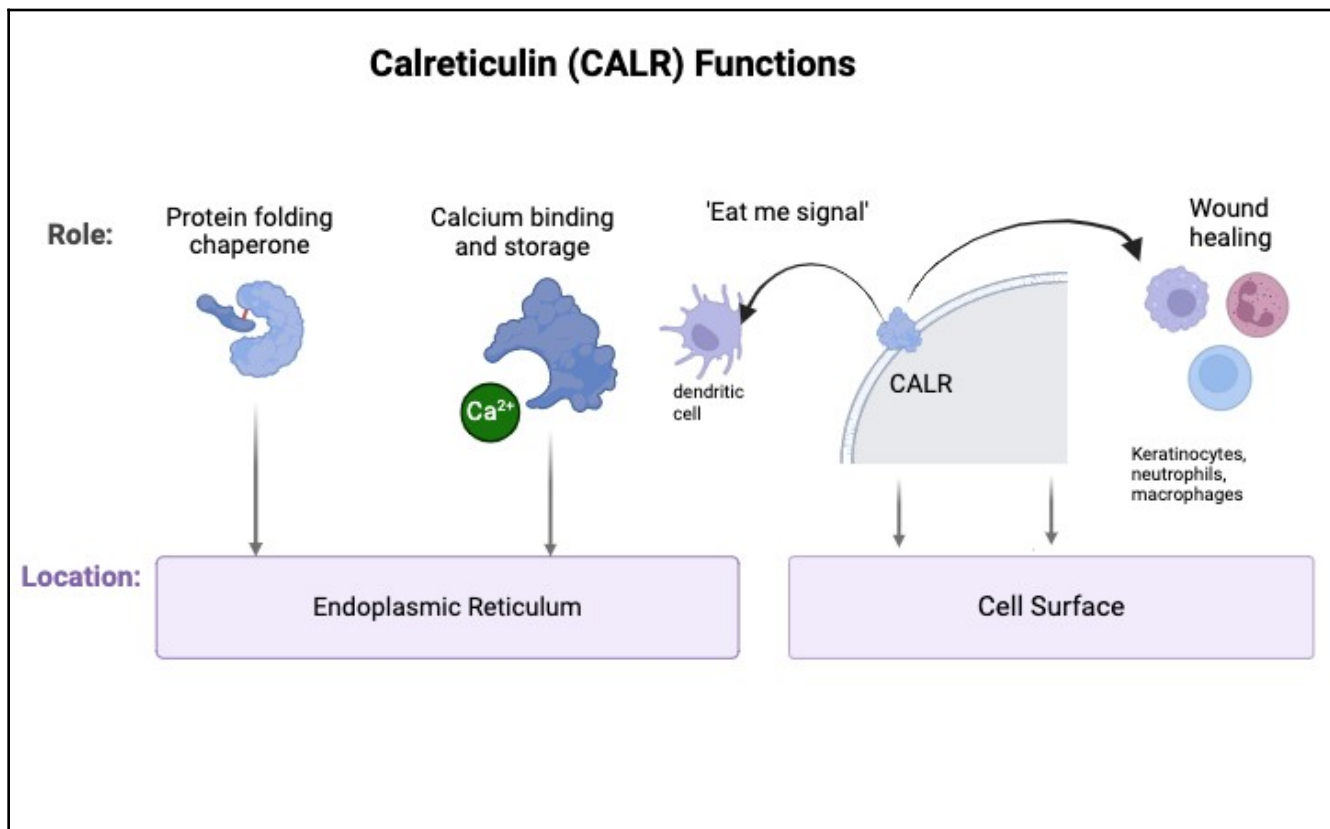


Figure 5. Diagram showing major roles of calreticulin in the cell: facilitating protein folding and maturation, calcium binding and retention, wound healing, then sending an 'eat me' signal to nearby dendritic cells as part of apoptosis [75-77]. Image generated by Biorender.

Another calcium-related process is the unfolded protein response (UPR), or the prevention of movement of misfolded or immature proteins from the ER to the Golgi apparatus. In tandem with calnexin, CALR refolds these proteins by glycosylation as a quality control measure: if this does not rectify the improper folding, the protein in question is then targeted by ERAD, or ER-associated degradation. CALR does not have the capacity to recognize misfolded proteins on its own: UDP-glucose and glycosyltransferases target misfolded glycoproteins and re-orient them in such a way that they almost perfectly regain their native conformations, thus facilitating binding by the CALR-calnexin

complex. Not all misfolded proteins interact this way with CALR inside the ER: hydrophobic proteins have different chaperones, such as the HSP70 family, that have more compatible domains to ensure correct folding [79].

In addition to reacting to stress inside the ER when misfolded proteins begin to aggregate, CALR also plays a role when ER stress surpasses the threshold required to initiate autophagy. Although ERAD is usually sufficient to contain misfolded proteins that cannot be managed by CALR or similar chaperone proteins, this does not always alleviate any present or potential ER stress. Autophagy is another type of quality control where cellular components are degraded by the lysosome, and CALR synthesis is stimulated by aggregation of misfolded proteins inside the ER, which in turn initiates CALR interactions with the microtubule-associated protein 1A/1B light chain 3 (LC3). There is a specific motif (LIR) in human CALR residues 200-204 that binds to LC3, and loss of this motif increases ER stress due to the lack of LC3 activity. Lysosomes are recruited in turn by LC3, which is then re-arranged and elongated to form an autophagosome to dispose of the aggregated pernicious proteins and provide a negative feedback for the ER stress [80].

Together with its role in facilitating autophagy, CALR also plays a role in immunogenic cell death in response to ER stress by relocating to the cell surface with ERp57 through SNAREs, or exocytic vesicle related proteins. Pancreatic ER kinase (PERK) responds to this stress by phosphorylating the eukaryotic translation inhibition factor, which leads to the cleavage and activation of, respectively, the pro-apoptotic caspase 8 and Bax, followed by CALR/ERp57 translocation. Intracellular calcium depletion has also been associated with this process as it removes activity that would otherwise retain CALR inside the ER. CALR is N-linked glycosylated at two aspartic acids residues 162 and 195, which allows it to move away from its typical location inside the ER, through the

Golgi apparatus, and then further on to the plasma membrane, with none found in the cytosol [73, 81]. As it arrives at the cell surface, CALR shifts to functioning as a damage-associated molecular pattern (DAMP) that is recognized by pattern recognition receptors (PRRs) on dendritic cells, in particular CD91. These dendritic cells then recruit cytotoxic T cells that engulf and degrade the CALR-exposing cell as part of the immune response [82, 83].

A highlight of the functional versatility of CALR is its ability to stimulate wound healing in addition to, and partly as a consequence of, cell death. One benefit of these cells dying by phagocytosis is that they may be replaced by new ones, and CALR helps fulfill this need. Utilizing its cell-surface relocation capacity as well as the ability to recruit keratinocytes to form focal adhesions, CALR is directly correlated with wound healing in mammalian tissues. In mice, pigs, and humans, CALR triggers migration and proliferation of keratinocytes, fibroblasts, and macrophages to the wounded area as a surface cytokine such that the collagen used to build new cells inside the tissue is evenly distributed and also that decaying damaged cells are destroyed. Even though it is at the surface, it still employs calcium to perform these activities as well, illustrating the dependency of CALR on this metal ion. Although most cells are activated by surface CALR, the protein is secreted from neutrophils to activate other neutrophils nearby to reduce the oxidative stress damaging the original tissue [84]. Since many CALR functions are dependent on its calcium capturing ability, and also because most of this ability is concentrated in the C-terminus, the somatic mutations that occur there have a significant impact on the protein's function and have a correlation with several diseases [85].

1.6.2) CALR Mutations and MPN Disease Progression

CALR mutations are observed in PMF and ET patients, but not in PV patients. These mutations are mostly indels, with two main types: del52 (also referred to as type 1) and ins5 (type 2). The first is the result of a 52-bp deletion and the second stems from a 5-bp insertion: they both occur at the C-terminus, but at different locations: leucine 367 for del52 and lysine 387 for ins5. One functional consequence of the ins5 mutation happening at a further downstream location is that some of the acidic residues required for calcium retention are still retained (see Table 2 for more details). There was a striking mutual exclusivity between *CALR* and *MPL/JAK2* mutations observed in a 2013 study conducted by Nangalia *et al* that was not seen in other studies [37]. As seen in Table 2, the mutations occur in exon 9, and consequences are the loss of both negatively charged amino acids and the KDEL retention sequence [37, 38]. *CALR* wild-type functions as a chaperone in the ER, and both type 1 and type 2 mutations are detrimental to this role, resulting in an increase of improperly folded proteins, although this only occurred in patients who were homozygous for either mutation, and those patients are a tiny minority. There have been other types of mutations seen in other patients, but many of them have been only observed in a single patient and are not functionally different from the two main types [39, 40]. Given their extreme rarity and functional similarities, they will not be elaborated on in any further detail.

Table 2. Comparison of *CALR* wild type sequence with those of the two main types of mutants, type 1(del52) and type 2(ins5). New C-terminal sequence for mutants highlighted in red. Sequence highlighted in blue represents the negatively charged sequence present in the wild-type that is conserved in the ins5 sequence. Leucine 367 (red) and Lysine 385 (light blue) have been highlighted and increased in size to show the points where the 52-bp deletion and 5-bp insertion occur, respectively.

CALR variant	Sequence
WT(wild-type)	MLLSVPLLLGLLGLAVAEPVYFKEQFLDGDGWTSRWIESKHKSDFGKFVL SSGKFYGDDEEKDKGLQTSQDARFYALSASFEPFSNKGQTLVVQFTVKHEQNI DCGGGYVKLFPNSLDQTDHMGDSEYNIMFGPDICGPGTKKVHVIFNYKGKN VLINKDIRCKDDEFTHLYTLIVRPDNTYEVKIDNSQVESGSLEDDWDFLPPK KIKDPDASKPEDWDERAKIDDPDTSKPEDWDKPEHIPDPDAKKPEDWDEEM DGEWEPPIQNPEYKGEWKPRQIDNPDYKGTWIHPEIDNPEYSPDPSIAYD NFGVLGLDLWQVKSQTIFDNFLITNDEAYAEFEGNETWGVTKAAEQMKD KQDEEQRLKEEEEDKKRKEEEEAEDKEDDEDKDEDEDEEDKEEDEED VPGQAKDEL
Del52 (type 1)	MLLSVPLLLGLLGLAVAEPVYFKEQFLDGDGWTSRWIESKHKSDFGKFVL SSGKFYGDDEEKDKGLQTSQDARFYALSASFEPFSNKGQTLVVQFTVKHEQNI DCGGGYVKLFPNSLDQTDHMGDSEYNIMFGPDICGPGTKKVHVIFNYKGKN VLINKDIRCKDDEFTHLYTLIVRPDNTYEVKIDNSQVESGSLEDDWDFLPPK KIKDPDASKPEDWDERAKIDDPDTSKPEDWDKPEHIPDPDAKKPEDWDEEM DGEWEPPIQNPEYKGEWKPRQIDNPDYKGTWIHPEIDNPEYSPDPSIAYD NFGVLGLDLWQVKSQTIFDNFLITNDEAYAEFEGNETWGVTKAAEQMKD KQDEEQRTRRMMRTKMRRMRMRRTRRKMRRKMSPARPTSCREACLQGWTEA
Ins5 (type 2)	MLLSVPLLLGLLGLAVAEPVYFKEQFLDGDGWTSRWIESKHKSDFGKFVL SSGKFYGDDEEKDKGLQTSQDARFYALSASFEPFSNKGQTLVVQFTVKHEQNI DCGGGYVKLFPNSLDQTDHMGDSEYNIMFGPDICGPGTKKVHVIFNYKGKN VLINKDIRCKDDEFTHLYTLIVRPDNTYEVKIDNSQVESGSLEDDWDFLPPK KIKDPDASKPEDWDERAKIDDPDTSKPEDWDKPEHIPDPDAKKPEDWDEEM DGEWEPPIQNPEYKGEWKPRQIDNPDYKGTWIHPEIDNPEYSPDPSIAYD NFGVLGLDLWQVKSQTIFDNFLITNDEAYAEFEGNETWGVTKAAEQMKD KQDEEQRNCLKEEEDDKRKEEEEAEDRRMMRTKMRRMRMRRTRRKMRR KMSPARPTSCREACLQGWTEA

The two types of mutants behave similarly in that both are secreted outside of the cell, but at different rates: type 1 is secreted at higher rates than type 2 even at a lower allele burden [41]. The increase in secretion does affect the interaction with immune cells: whereas the wild type *CALR* is moderately secreted, allowing for dendritic cells to be matured by activating T cells, this does not happen with mutant *CALR* which over-saturates the receptors on the dendritic cells before they can mature. Since they are unable to mature properly, this renders them ineffective in killing the cells that secrete the mutant, allowing them to continue to proliferate [44].

Both mutant CALR types bind to MPL inside the cell and are secreted together outside the cell to bind to nearby cells for initiating thrombocythemia when mutated, and mutations in both of these genes are frequently found in ET and MF patients [68]. Simultaneously, MPL is trafficked to the cell surface by mutant CALR, which in turn results in increased ERK phosphorylation and cell proliferation characteristic in MPNs [69]. This only occurs in cell with mutant calreticulin, as the wild-type is incapable of binding to MPL because of the incompatibility of its C-terminus with MPL and its inability to homo-dimerize, both of which are needed for MPL translocation and subsequent activation [70].

The consequences of the *CALR* mutations versus the other two types of driver mutations, *JAK2* and *MPL*, is the marked decrease in thrombocytopenia, possibly due to the mutant gaining the ability to bind to and sequester MPL [45]. It is also considered a driver mutation because it brings a specific profile of MPNs that is distinct from either *JAK2* or *MPL* and that it is usually the first mutation detected during the early phase of abnormal HSC proliferation, followed later by mutations in passenger genes [46]. As mentioned in the previous section, another main role of CALR is binding to calcium and regulating its intracellular levels, and it is able to do this thanks to its negatively charged regions with high amounts of glutamic and aspartic acids in the C-terminal domain. However, as a consequence of the frameshift mutations, the acidic nature is lost and replaced with positively charged lysines and arginines which repel the calcium due to sharing a positive charge. Due to the loss of calcium-CALR interactions, NFAT cannot localize to the nucleus and negatively regulate myeloid proliferation, which is why PMF patients, with higher myeloid counts, always have the *CALR* mutations [47]. This calcium regulation system is known as store-operated calcium (2+) entry (SOCE) and consists of a complex composed of the calcium ions, wild-type CALR, and stromal interaction

molecule 1 (STIM1). Since most MPN patients are just heterozygous for mutant *CALR*, this system is significantly impaired rather than completely deleted. Roughly half of *CALR* mutant megakaryocytes (both type 1 and type 2) had significantly higher amounts of spontaneous calcium spikes compared to healthy controls, and this positively correlated with higher levels of megakaryocyte proliferation. The type 1 cells had higher levels of SOCE dysfunction, most likely attributed to its higher secretion levels compared to type 2. The correlation between SOCE dysfunction and mutant *CALR* was confirmed even further by the use of SOCE inhibitor BTP-2 that abated proliferation of both wild-type and mutant *CALR* [48].

The second major effect of the mutations on *CALR* is their ability to bind to MPL but this only occurs in the presence of JAK2. Specifically, the asparagine 117 on the MPL extracellular domain glycan binding site has been shown to be crucial for activation by mutant *CALR* as its removal significantly reduced STAT5 transcription. The MPL that is sequestered to the cell surface is immature, and is also less sensitive to cytokine signaling: in the absence of Tpo and with the addition of either *CALR* del52 or ins5, STAT5 transcriptional activity correlating with greater cell proliferation, was twice as high as that of the wild-type [50, 66].

Additionally, mutant *CALR* also binds to the transferrin receptor (TfR) as part of the activation mechanism for MPL: TfR stabilizes and significantly extends the half-life of mutant *CALR* (either del52 or ins5). *CALR* WT does not interact with TfR in this way, likely due to its KDEL retention signal and relative lack of secretion capacity. Mutant *CALR* (especially del52) is secreted from one cell while bound to the soluble domain of TfR (sTfR) and moves through the plasma to bind to MPL receptors on other cells to activate the JAK-STAT pathway to initiate proliferation, leading to a positive feedback loop where these clones gradually dominate the hematopoietic landscape. These cells

are characterized by MPL receptors with immune N-linked polysaccharides, which are exclusively capable of interacting with this CALR-sTfR complex. The P-domain in mutant CALR binds to these polysaccharides, which is still the same as the one in CALR WT, but only the latter has the opportunity to perform this role due to the frameshift mutation occurring towards the C-terminus [90].

1.6.3) Mutant CALR and the Unfolded Protein Response

Apart from MPL and TfR, mutant CALR also has a repertoire of other proteins that it has similar types of interactions with, or the interactome, in several organelles, including the cytoplasm, nucleus, and ER. Two proteins associated with the UPR, BiP and Grp75, bind to mutant CALR as well as WT CALR [99]. However, the impact on the UPR depends on the type of mutant CALR, as the C-terminus for ins5 is much more similar to WT than del52 (see Table 2) in that ins5 still has some of the acidic residues that can bind to calcium ions. The significant drop in SOCE activity results in activation of the UPR by splicing mRNA and activating of XBP1 (Figure 6), whose target genes start to alleviate ER stress. Consequently, this leads to up-regulation of the anti-apoptotic Bcl-2 and clonal expansion of CALR del52 cells: this effect was not seen in ins5 or WT CALR cells [100].

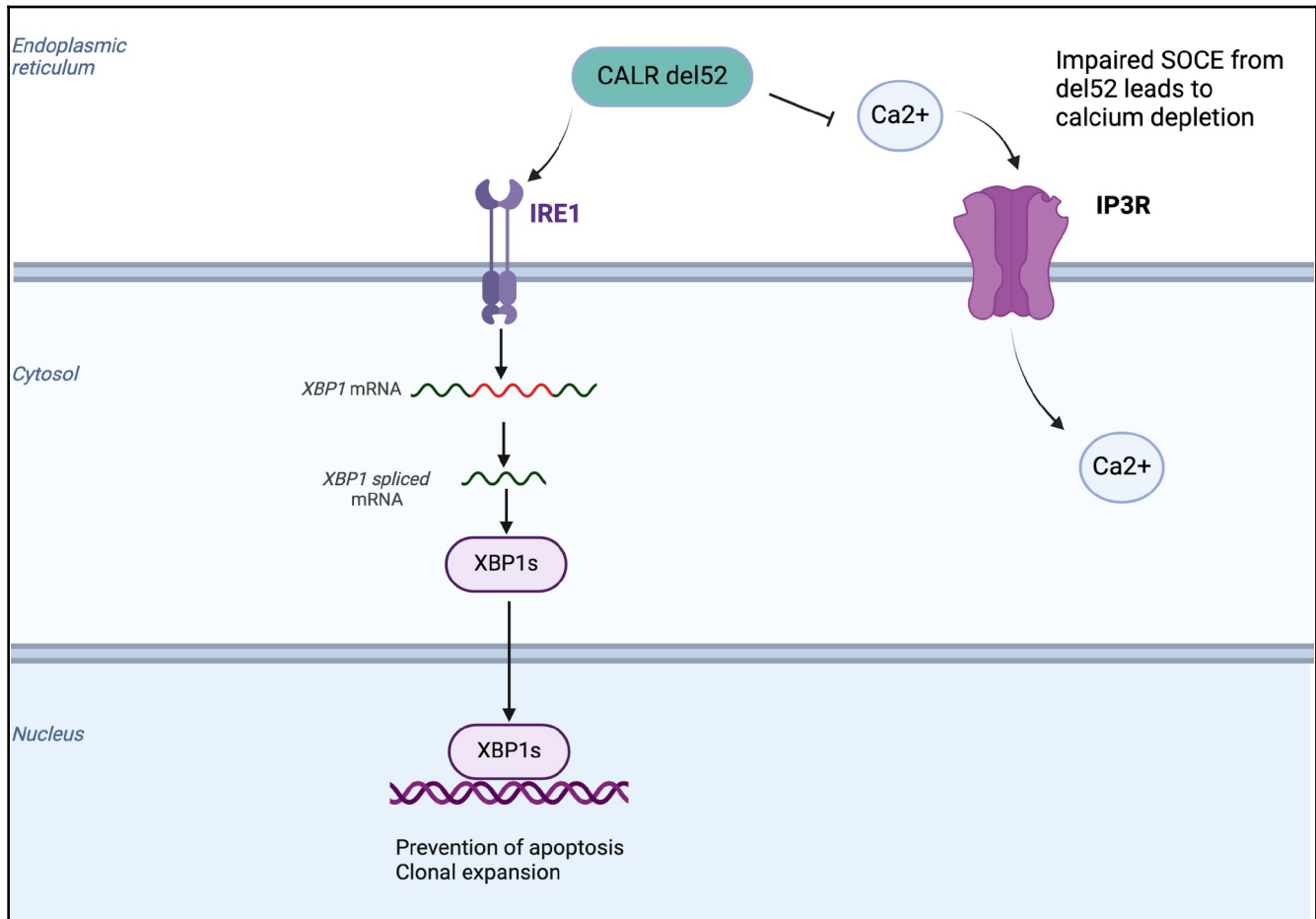


Figure 6. Display of CALR del52 contributing to MPN development by impairment of SOCE function, which in turn contributes to XBP1 activation and clonal expansion [100]. Image generated by Biorender.

Another consequence of these frameshift mutations that contributes to the heightened UPR is the homo-multimerization of mutant CALR as well as calcium loss, both of which contribute to aggregation of misfolded proteins, partially due to myeloperoxidase (MPO) deficiency. MPO is a myeloid-specific enzyme that catabolizes hydrogen peroxide as part of the immune response and is typically glycosylated and matured by WT CALR in order to function, but in cells with these frameshift mutations, it is targeted for degradation by the proteasome due to its immaturity. MHC class I glycoproteins are also affected in the same manner and thus this type of immune response is more muted in CALR mutant cells. There are both global structural alterations in glycoprotein content and localized changes in calcium-related proteins folded by CALR when either of these frameshift mutations occur, although the effect is much stronger when the mutation is homozygous. Most MPN patients are heterozygotes, so the overall decrease in glycoprotein stability is not as drastic as when the mutation is homozygous. Part of the MPO loss stems from mutant CALR secretion outside the ER, but even when CALR is inside the ER, it cannot bind to ERp57 properly, resulting in improper folding and later degradation of substrates like MPO [88].

Going beyond the relationship with MPL, studies in K562 cells that do not endogenously express the receptor have shown that mutant CALR causes another problematic situation for the cell besides an impaired ER stress response, particularly accumulated ROS. Superoxide dismutase, the enzyme that converts superoxide to hydrogen peroxide and oxygen, is significantly down-regulated in the presence of either mutant compared to CALR WT. Unlike the UPR response, both del52 and ins5 show an impaired capacity to remedy increasing oxidative stress inside the cell. As a result of this damage, DNA is also significantly affected due to silencing of the oxidative stress repair gene *OXR1* that occurs in the presence of mutant CALR, but not WT. Not only does this result in higher levels of

apoptosis, but in the surviving cells, there are also much higher levels of genomic instability and more frequent mutations. However, there are more apoptotic cells, so the net effect is increased cell death. The clinical implications of this are that, at least in MF, CALR mutant patients have better outcomes compared to those with JAK2 V617F, mainly because the CALR mutant blocks further mutations that increase the likelihood of progression to sAML by preventing the conditions necessary for them to occur. This mechanism is partially why most MPN patients do not undergo leukemic transformation [101, 102].

1.7) Aims and Objectives

Given the importance of the effects of the CALR frameshift mutations that occur in MPN patients and how they disrupt the crucial functions of this ubiquitous chaperone, the aim of this thesis is to explore why and how this novel C-terminus is able to facilitate MPN development. The current consensus is that this is mostly attributed to the loss of the KDEL anchor motif, but this explanation does not account for the acquisition of secretory capabilities that either type of CALR mutant possesses. Furthermore, there are no studies to our knowledge that characterize the properties of this novel C-terminal sequence in depth and answer the questions as to why and how mutant CALR is secreted outside its typical residence in the ER. However, studies on membrane proteins [103, 104] have found that they employ motifs rich in arginines, similar to mutant CALR (especially del52, see Table 1 for more details), and also based on the central theme in biology of structure determining function, it is likely that these motifs also facilitate mutant CALR secretion.

Preliminary data by Chachoua et al. (unpublished) has also shown that modification of the mutant C-terminus through either truncation or mutation affects its subcellular localization, and mutant

secretion and STAT5 activation is inversely proportional to arginine content. This has been shown in Western blots and luciferase-based STAT5 activity assays where stop codons that removed more of the novel C-terminus resulted in larger drops in secretion and STAT5 activity. Furthermore, the addition of the KDEL anchor motif to the del52 sequence did not decrease STAT5 activity, suggesting that this modified del52 still has some secretory capacity thanks to these arginine-rich motifs that are able to counteract the retention tendency of KDEL.

Moreover, other research groups have shown that deletion of the C-terminus in mutant CALR results in more CALR expression inside the cell. However, their interpretation was that truncated mutants have higher steady-state protein expression [66, 105]. Previously, two types of CALR mutants were discussed, but in actuality these are more like categories rather than distinct sequences. Studies on CALR mutations show some variety in the sequences, but the common denominator is a C-terminus with arginine-rich motifs (see Tables 3a and 3b, residues highlighted in green). Some patients have extra motifs that also contain arginines or lysines (residues highlighted in black), but all of these sequences functionally converge and result in a version of CALR that does not remain in the ER and instead is secreted outside the cell. For the sake of simplicity, only the consensus sequences will be referred to from now on (see Tables 2 and 3c).

Table 3b. List of documented CALR ins5-like (also known as type 2) sequences according to the Mayo Clinic classification scheme. Residues highlighted in green are shared between patients, while those highlighted in black are unique to individual patients. Number of affected nucleotides and location along the protein sequence are also shown. Data adapted from Lasho et al [57].

Type 2 (ins5)	c.1111_1147del (c.1109_1145del)	E371fs*47	AAEQMKDKQDEEQRLKEERTRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1113A>C;1122_1140del	E371D+K375fs*49	AAEQMKDKQDEEQRLKEEEDK RQRTRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1114_1144del	E372fs*48	AAEQMKDKQDEEQRLKEEEQ RTRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1116_1146del	D373fs*47	AAEQMKDKQDEEQRLKEEEER TRRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1118_1145delinsCGTTTA	D373fs*50	AAEQMKDKQDEEQRLKEEE AFKTRRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1120_1138del	K374fs*50	AAEQMKDKQDEEQRLKEEED RRQRTRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1121_1148del	K374fs*47	AAEQMKDKQDEEQRLKEEED TRRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1124_1133del	K375fs*52	AAEQMKDKQDEEQRLKEEED KRRRRQRTRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1124_1142del (c.1122_1141delinsA)	K375fs*49	AAEQMKDKQDEEQRLKEEEDK RQRTRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1125_1146del	K377fs*46	AAEQMKDKQDEEQRLKEEEDK KRTRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1131_1152del	E378fs*45	AAEQMKDKQDEEQRLKEEEDK KRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1135_1147del	E379fs*47	AAEQMKDKQDEEQRLKEEEDK KRKERTRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1140del	E381fs*49	AAEQMKDKQDEEQRLKEEEDK KRKEEERQRTRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1143_1154delinsTCCTTGTC	E381fs*48	AAEQMKDKQDEEQRLKEEEDK KRKEEERMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1147_1156del/insTGTCGG	E383fs*46	AAEQMKDKQDEEQRLKEEEDK KRKEEEEACRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1151A>G;1154_1155insTTGTC	D384G;K385fs*47	AAEQMKDKQDEEQRLKEEEDK KRKEEEEAGNCRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1153_1154delAA/insCCTTGTC	K385fs*47	AAEQMKDKQDEEQRLKEEEDK KRKEEEEAEDPCRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1153_1154delinsTCTTGTC	K385fs*47	AAEQMKDKQDEEQRLKEEEDK KRKEEEEAEDSCRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1153_1154delinsTGTC	K385fs*46	AAEQMKDKQDEEQRLKEEEDK KRKEEEEAEDCRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1154_1155insTTGTC **	K385fs*47	AAEQMKDKQDEEQRLKEEEDK KRKEEEEAEDNCRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1154_delA/insCTTGTC	K385fs*47	AAEQMKDKQDEEQRLKEEEDK KRKEEEEAEDTCRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1154_delA/insTCTTGTC (c.1154_delinsTTTGTC)	K385fs*47	AAEQMKDKQDEEQRLKEEEDK KRKEEEEAEDICRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1154_delA/insTGTGTC	K385fs*47	AAEQMKDKQDEEQRLKEEEDK KRKEEEEAEDMCRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA
Type 2 (ins5)	c.1155_1156insTGTCG (c.1154_1155insATGTC)	E386fs*46	AAEQMKDKQDEEQRLKEEEDK KRKEEEEAEDCRRMMRTKM MRMRMRTRRKMRKMSPARPRTSCREACLQGWTEA

Table 3c. Full consensus sequences of CALR del52 (top) and WT CALR (bottom). Highlighted motifs represent proposed motifs that contribute to secretion to mutant CALR (top) and residues that will be mutated to motifs similar to what is present in del52 (bottom).

CALR sequence	del52	MLLSVPLLLGLGLAVAEPVYFKEQFLDGDGWTSRWIESKHKSDFGKFVLSGKFYGDDEKDKGLQTSQDARFYALSA SFEPFSNKGQTLVVQFTVKHEQNIDCGGGYVKLFPNSLDQTDHMGDSEYNIMFGPDICGPGTKKVHVIFNYKGKNVLINK DIRCKDDEFTHLYTLIVRPDNTYEVKIDNSQVESGSLEDDWDFLPPKKIKDPDASKPEDWDERAKIDDPDTSKPEDWDKPE HIPDPDAKKPEDWDEEMDGEWPPVIQNPEYKGEWKPRQIDNPYKGTWIHPEIDNPEYSPDPSIAYDNFGVLGLDLWQ VKSGTIFDNFLITNDEAYAEFGNETWGVTKAAEKQMKDKQDEEQRT RRRM MR TKMR MRMR RRTR RRK MRRK MSPARPR TSCREACLQGWTEA
CALR WT		MLLSVPLLLGLGLAVAEPVYFKEQFLDGDGWTSRWIESKHKSDFGKFVLSGKFYGDDEKDKGLQTSQDARFYALSASF EPFSNKGQTLVVQFTVKHEQNIDCGGGYVKLFPNSLDQTDHMGDSEYNIMFGPDICGPGTKKVHVIFNYKGKNVLINKDIR CKDDEFTHLYTLIVRPDNTYEVKIDNSQVESGSLEDDWDFLPPKKIKDPDASKPEDWDERAKIDDPDTSKPEDWDKPEHIPD PDAKKPEDWDEEMDGEWPPVIQNPEYKGEWKPRQIDNPYKGTWIHPEIDNPEYSPDPSIAYDNFGVLGLDLWQVKSG TIFDNFLITNDEAYAEFGNETWGVTKAAEKQMKDKQDEEQRLKEEEEDK KRK EEEEAE D KEDDE DKDE DEDE EDKE EDEEDVPGQAKDEL

The objectives of this project are as follows: to predict the roles of possible secretion motifs found in the novel C-terminal sequence of both del52 and ins5 CALR mutants using *in silico* analyses, and using tools and methods available online or on GitHub repositories to predict the effects of losing or gaining these motifs in mutant and WT CALR, respectively. The general consensus about CALR is that it is able to be retained in the ER thanks to the KDEL anchor sequence [107], which is then lost in ins5 or del52 mutants. However, we posit that the loss of KDEL is not sufficient for the subsequent gain of secretion ability, but what is also required is the acquisition of positively charged residues resulting from the frameshift mutations. We will therefore focus on *in silico* analyses concerning these motifs and how their losses in del52 and gains in CALR WT through substituting acidic residues will affect predicted secretion, which is crucial to mechanisms involving binding to MPL and proliferation of MPNs [90, 105].

CHAPTER 2

MATERIALS AND METHODS

MATERIALS

2.1.1) Reagents for competent *E. coli* growth and storage

Reagent Name	Product Code	Manufacturer
Agar	05039-500G	Sigma Aldrich (St. Louis, Missouri, USA)
Ampicillin sodium salt	13399.02	SERVA (Heidelberg, Germany)
Sodium chloride	969.036	ISOLAB (Eschau, Germany)
Yeast extract	YEX401.500	Biobasic Canada (Markham, Ontario, Canada)
Tryptone	800-158-LG	MultiCell (Saint Jean-Baptiste, Quebec, Canada)
Calcium chloride dihydrate (Inoue buffer)	C3306-500G	Sigma Aldrich (St. Louis, Missouri, USA)
Potassium chloride (Inoue buffer)	1.04936.1000	Millipore (Burlington, Massachusetts, USA)
PIPES Buffer Grade	A1079.1000	AppliChem (Darmstadt, Germany)
Manganese chloride	1.05917.1000	Millipore (Burlington, Massachusetts, USA)
Glycerol (87%)	39788.01	SERVA (Heidelberg, Germany)

2.1.2) HEK293 cell media reagents

Reagent Name	Product Code	Manufacturer
Fetal Bovine Serum	04-007-1A	Biological Industries (Beit Haemek, Israel)
100x Penicillin/Streptomycin	PS-B	Capricorn Scientific

solution		(Ebsdorfergrund, Germany)
0.25% Trypsin, 2.21 mM EDTA	325-043-EL	MultiCell (Saint Jean-Baptiste, Quebec, Canada)
Dulbecco's PBS	BDL-001-500ML	Serana (Brandenburg, Germany)
DMEM High Glucose with L-Glutamine without Sodium Pyruvate	SLD-524-500	Serox (Mannheim, Germany)
DMSO	D0231	Biobasic Canada (Markham, Ontario, Canada)
100 cm cell culture dishes	L231100	Serox (Mannheim, Germany)
Potassium chloride	1.04936.1000	Millipore (Burlington, MA, USA)
Sodium hydrogen phosphate	1.06580.1000	Millipore (Burlington, MA, USA)
HEPES, free acid	0511-1KG	AMRESCO (Solon, Ohio, USA)
Dextrose	16325	Riedel de Haën (Seelze, Germany)
Calcium chloride dihydrate	C3306-500G	Sigma Aldrich (St. Louis, Missouri, USA)

All cells maintained at 37°C in a Nu-AIRE DH Autoflow incubator (Leeds, United Kingdom).

2.1.3) List of mutagenesis primers

Mutagenesis	Forward Primer	Reverse Primer	Forward T _m	Reverse T _m
CALR del52 A394*	agctcgtccttgccctctacggg acatcttcctc	gaggaagatgtccccgtagaggccaaggac gagct	78.24°C	78.24°C
CALR del52 M387*	gggacatcttctctctacttctc cttgctctcc	ggaggacaaggaggaagtagaggaggaag atgtccc	79.97°C	79.97°C
CALR del52 M377*	gtcctctcatcctctctacatcat cttgctctcat	atgaggacaaagatgaggttagaggaggatg aggaggac	78.97°C	78.97°C
CALR del52 M371*	catcctcatcttgctctacatcct ccttgctctctgc	gcaggagacaaggaggatgtagaggacaa agatgaggatg	80.12°C	80.12°C
CALR del exon 9	atcctccttgctctactcctcgtc ctgttg	caaacaggacgaggagtagaggacaagga ggat	79.14°C	79.14°C

CALR del52 All R	catcttcctccttcctcctcctcctcctcctcctcctcctcatcttgt	acaaagatgaggaggaggaggaggaggaggaggagaaggaggaagatg	79.88°C	79.88°C
CALR del52 All M	ccggggacatcttcctcctcatcttcctcatcatcatcatcatcatcatcatc	gatgatgaggacaaagatgaggatgatgatgatgatgatgatgatgaagatgaggaggaa	79.30°C	79.30°C

2.1.4) List of Western blot reagents (excluding buffers)

Reagent Name	Product Code	Manufacturer
Anti-Calreticulin Rabbit Monoclonal Antibody [ARC51269]	a305767	Antibodies.com (Cambridge, United Kingdom)
Süttozu (milk powder)	TR 35-0067	Pınar (İzmir, Türkiye)
Anti-Beta Actin Rabbit Monoclonal Antibody	M01263	BosterBio (Pleasanton, California, USA)
Protease Inhibitor Mix M	SE3910201	SERVA (Heidelberg, Germany) (Heidelberg, Germany)
Opti-Protein Marker, 10kDa-175kDa, 500 µl/100 loads	G252	Abmgood (Richmond, BC, Canada)
ClearBand Laemmli Sample Buffer (4x), 15ml	LSB-2x	Ecotech Biotech (Yakutiye, Türkiye)
Glycine	927.032.1000	ISOLAB (Eschau, Germany) (Eschau, Germany)
Acrylamide/Bis Solution 37.5:1 (30% w/v), 2.6% C	SE1608801	SERVA (Heidelberg, Germany) (Heidelberg, Germany)
Albumin, heat shock isolation, pH 7.0	AD0023	Biobasic Canada (Markham, Ontario, Canada)
0.22 µM ClearBand PVDF membrane	PVDF022	Ecotech (Yakutiye, Türkiye)
HRP Conjugated Goat Anti-Rabbit Secondary Antibody	BA1054	BosterBio (Pleasanton, California, USA)
Whatman Gel Blotting Paper 70×100 mm	WHA10427806	Whatman (Maidstone, United Kingdom)

2.1.5) List of Western blot buffers with individual ingredients

Reagent Name	Buffer Type	Product Code	Manufacturer
NaCl	NP-40 lysis	969.036	ISOLAB (Eschau, Germany)
Nonidet P-40	NP-40 lysis	74385-1L	Sigma Aldrich (St. Louis, Missouri, USA)
Tris	NP-40 lysis	37190.2	SERVA (Heidelberg, Germany)
Ponceau S	Ponceau staining	A1405.0025	AppliChem (Darmstadt, Germany)
Acetic acid	Ponceau staining	9.010.132.5	ISOLAB (Eschau, Germany)
SDS	Running buffer	8.22050.1000	Merck (Rahway, NJ, USA)
Glycine	Running buffer	927.032.1000	ISOLAB (Eschau, Germany)
Tris	Running buffer	37190.2	SERVA (Heidelberg, Germany)
Tris	Transfer buffer	37190.2	SERVA (Heidelberg, Germany)
Glycine	Transfer buffer	927.032.1000	ISOLAB (Eschau, Germany)
Tris-HCl	TBS	37192.01	SERVA (Heidelberg, Germany)
NaCl	TBS	969.036	ISOLAB (Eschau, Germany)
Tween-20	TBS-T (add to TBS separately to make TBST)	39796.51	SERVA (Heidelberg, Germany)
Methanol	Transfer buffer	24229-2.5L-R	Sigma Aldrich (St. Louis, Missouri, USA)

2.1.6) Mutagenesis, plasmid miniprep kits, DNA measurements, and agarose gels

For site-directed mutagenesis, we used the Phusion High Fidelity Kit from New England Biolabs (product code E0553S, Ipswich, Massachusetts, USA), and although this kit does include dNTPs, we also used a separate dNTP mix from Ampliqon (product code A502004, Odense, Denmark) as part of the reaction mix as well. Following the initial mutagenesis step, we used DpnI to digest the remaining original plasmid (product code R0176, Ipswich, Massachusetts, USA). The incubator shaker and benchtop incubator used for bacterial transformation were the Innova 4300 from New Brunswick Scientific (Edison, New Jersey, USA) and BD56 from Binder GmbH (Tuttlingen, Germany), respectively. Additionally, the UV-Vis spectrophotometer used for OD600 readings was the SOIF model from Nukleon Lab (Ankara, Türkiye). Three different types of miniprep kits were used depending on availability and price: the GeneJET Plasmid Miniprep Kit from Thermo Fisher Scientific (product code K0502, Waltham, Massachusetts, USA), the Spin Miniprep Kit from Qiagen (product code 27104, Hilden, Germany), or the New England Biolabs Monarch Plasmid Miniprep Kit (product code T1010, Ipswich, Massachusetts, USA). For agarose gels, agarose from BioShop Canada (product code AGA001.100, Burlington, Ontario, Canada) was used at 1% w/v, with DNA Stain G from SERVA Electrophoresis GmbH (product code 39803.01, Heidelberg, Germany) acting as a dye. Milli-Q water was used to dilute the samples along with 3x loading dye (product code 7222, Cell Signaling Technology, Boston, Massachusetts, USA). Samples were either visualized at 450 nm on either the Amersham Imager 600 (Amersham, United Kingdom) or the Vilber-Lourmat UV transilluminator (model unknown, Marne-la-Vallée, France). The TAE (Tris base-acetic acid-EDTA) buffer was initially prepared at 50x concentration, then diluted to 0.5x concentration from the following: Tris base

(product code 37190.2, SERVA, Heidelberg, Germany), acetic acid (product code 9.010.132.5, ISOLAB, Eschau, Germany), and EDTA (product code 1.08421.1000, Darmstadt, Germany).

2.1.7) List of plasmids used for mutageneses

pMSCV-IRES-mCherry plasmids containing an ampicillin resistance gene and either *CALR* del52, WT *CALR*, or an empty vector were graciously provided by Stefan Constantinescu's group at Louvain, Belgium. All three of these plasmids were approximately 7,000 bp in length and used the IRES (internal ribosome entry site) as a means of separating the *CALR* from the mCherry fluorescent marker such that only the protein was translated.

METHODS

2.2.1) DeepLoc sorting signals prediction

The first method of analysis was using DeepLoc [110], which provided predicted sorting signal importance by amino acid throughout the input sequence, although there is no distinction between different types of signals (for example, nuclear localization signals vs mitochondrial signaling). These scores were determined using a position-specific scoring matrix (PSSM), or assigning a score to each amino acid in the sequence based on its location and on similar motifs associated with these signals as seen across various proteins. In the case of sorting signal importance, these graphs were already generated by DeepLoc, using the Python package logomaker that generates a graph where the height of the amino acid letter corresponds to the predicted significance.

2.2.2) Eukaryotic Linear Motif (search engine tool

Following these results, the CALR del52 and WT sequences were entered into the ELM (Eukaryotic Linear Motifs) [108] website's search engine, which scans the sequences for likely or known motifs. These motifs are identified by their location in the sequence and their function similar to the DeepLoc methodology, typically between 3-15 amino acids in length and with predicted cellular localization(s) based on their function. Scores assigned to these motifs are based on the taxonomy of the species (in this case, *Homo sapiens*), position on the sequence, and experimental verification. Although not all mutant CALR sequences are exactly the same, there are consensus sequences for each type [106, 109], and our predicted motifs are thus based on these sequences.

2.2.3) Localization predictions with DeepLoc and MULocDeep

To predict the general localization of mutant CALR (especially del52) versus WT CALR, two prediction web tools were used. The first tool was DeepLoc [110], and as additional verification, the second was MULocDeep [111, 112]. Prediction scores for each organelle by DeepLoc rely on a transformer language model, a machine learning tool that factors in the location and type of each residue along the full sequence and compares it to a dataset from SwissProt containing thousands of proteins with empirically supported localizations, assigning probabilities based on sequence pattern recognition. MULocDeep works in a similar way to DeepLoc in that it assigns weighted values to each amino acid in the sequence based on its location, but the training data is derived from the UniProt database rather than SwissProt. Protein sequences were entered as input (see Tables 3a and 3b as an example), which then returned CSV files containing predicted localization scores for up to ten different organelles. CSV files were then uploaded to Anaconda Cloud notebooks and visualized using Python's

seaborn package and statistical analyses centered around ordinary least-squares regression were generated with the statsmodels.api package. Similar methods were used when trying different types of mutations, or when KDEL residues (to mimic the C-terminal anchor present on WT) were added sequentially to the C-terminus of CALR del52. For later analyses involving addition of KDEL residues to the CALR del52 sequence to see if there was a counter-effect to the pro-secretion properties of the proposed motifs, DeepLoc and MULocDeep were also used, with the Python package seaborn used to generate the heatmaps and regression plots.

2.2.4) Preparation of competent cells for bacterial transformation

Ultra-competent D5Hα E. coli cells were first prepared by spreading a previous batch onto a LB agar plate, which was incubated at 37 degrees Celsius overnight. Colonies were then streaked with a flame-sterilized pipette tips and added to fresh LB media to be incubated at 37 degrees Celsius for approximately 6 hours as a starter culture. Three flasks were then inoculated with various amounts of this starter culture, moved to the cold room to maintain a shaking temperature of approximately 18 degrees Celsius for a slower overnight growth. The next morning, OD600 readings were taken for each culture, with the sample with a OD600 of ~0.5 utilized and the other two flasks sterilized with bleach before discarding the broth. The selected flask's contents were then centrifuged at 4500 rpm for 10 minutes, and the cell pellets resuspended in Inoue transformation buffer (see section 2.1.1 for ingredients) with DMSO added later on. 50 and 100 uL aliquots were prepared with pre-cooled pipette tips and Eppendorf tubes, snap-frozen in liquid nitrogen, and moved to -80 degrees for long-term storage. The following day, aliquots were taken out of the freezer and transformed with 10 ng/uL and

100 ng/uL of a select plasmid to measure the transformation efficiency by counting the colonies on the plates and multiplying the number by the dilution factor of the plasmid.

2.2.5) Site-directed mutagenesis of CALR del52 truncations

The CALR del52 plasmid was graciously provided by Stefan Constantinescu at the Catholic University of Louvain in Belgium, and used as a template for mutagenesis: approximately 1 ng of plasmid was used for each reaction with the Phusion PCR kit provided by Thermo Fisher. The following PCR mix was used for each reaction:

Reagent	Amount
Phusion DNA Polymerase	0.1 uL
CALR del52 template DNA	~1 ng
5x HF Master Mix	2 uL
DMSO	0.5 uL
40 mM dNTP mix	0.1 uL
2 uM Forward Primer	0.5 uL
2 uM Reverse Primer	0.5 uL
Nuclease free water	Fill to 10 uL

The following PCR protocol was used for each reaction (20 cycles for denaturation, annealing, and extension):

Step	Time	Temperature (°C)
Initial denaturation	30 seconds	98
Denaturation	10 seconds	98
Annealing	20 seconds	Primer melting temperature – 5
Extension	9 minutes	72
Final extension	9 minutes	72

Hold	Indefinite	4
------	------------	---

After PCR was performed, the 10 uL reactions were treated with the DpnI restriction enzyme (1 uL) and then returned to the PCR machine for the following reaction to digest the remaining original CALR del52 plasmid:

Step	Time	Temperature (°C)
DpnI digestion	2-3 hours	37
Heat inactivation of DpnI	20 minutes	80
Hold	Indefinite	4

Once the plasmids were digested, 5 uL of this PCR reaction was transformed into 50 uL competent D5H α . These cells were incubated for 30 minutes on ice before a 75 second heat shock at 42 degrees Celsius. Approximately 1 mL of LB broth was added to these cells and the Eppendorf tubes were incubated at 37 degrees Celsius at 250 rpm for 40 minutes. After 40 minutes, the cells were centrifuged at 13,000 rpm for 1 minutes, with 900 uL of supernatant discarded. The cells were resuspended in the remaining LB broth before being spread onto LB plates with 50 ug/mL ampicillin, which were incubated overnight at 37 degrees. The following day, plates were inspected for colonies, and if colonies were found, they were selected for minipreps.

2.2.6) Miniprep procedures for mutagenesis colonies

Prior to use of these kits, a colony was randomly selected from each plate with a flame-sterilized pipette tip and placed in ~5 mL of LB broth with 50 ug/mL ampicillin to incubate overnight at 37 degrees Celsius at 225 rpm. The following day, the cultures were spun down at ~5000 rpm for 7

minutes, with the supernatant discarded afterwards. The bacterial pellets were then processed with any of the following kits: (note, all centrifugation done at 13,000 rpm, RT = room temperature)

Step	GeneJET Plasmid Miniprep Kit	QIAprep Spin Miniprep Kit	NEB Monarch Plasmid Miniprep Kit
Resuspension	250 uL Resuspension Buffer added, transfer resuspended pellet to Eppendorf	250 uL Buffer P1 added, transfer resuspended pellet to Eppendorf	200 uL Buffer B1 added, transfer resuspended pellet to Eppendorf
Lysis	250 uL Lysis Buffer added, tube inverted 4-6 times	250 uL Buffer P2 added, tube inverted 4-6 times	200 uL Buffer B2 added, tube inverted 4-6 times, incubate at RT for 1 min
Neutralization	350 uL Neutralization Buffer added, tube inverted 4-6 times, centrifuge for 5 min	350 uL Buffer N3 added, tube inverted 4-6 times, centrifuge for 10 min	400 uL Buffer B3 added, tube inverted 4-6 times, incubate for 2 min at RT, centrifuge for 5 min
Transfer supernatant to spin column, centrifuge for 1 min, discard flow through	Transfer supernatant to spin column, centrifuge for 1 min, discard flow through	Transfer supernatant to spin column, centrifuge for 1 min, discard flow through	Transfer supernatant to spin column, centrifuge for 1 min, discard flow through
Wash step #1	Add 500 uL Wash Buffer, centrifuge for 30 sec, discard flow-through	Add 500 uL Buffer PB, centrifuge for 30 sec, discard flow-through	Add 200 uL Wash Buffer 1, centrifuge for 1 min, discard flow-through
Wash step #2	Add 500 uL Wash Buffer, centrifuge for 30 sec, discard flow-through	Add 750 uL Buffer PE, centrifuge for 30 sec, discard flow-through	Add 400 uL Wash Buffer 2, centrifuge for 1 min, discard flow-through
Removal of remaining buffer	Centrifuge for 1 min, discard flow-through	Centrifuge for 1 min, discard flow-through	Centrifuge for 1 min, discard flow-through
Elution (transfer column to clean Eppendorf first)	Add 80 uL Elution Buffer, incubate at RT for 1-2 min, centrifuge for 2 min	Add 80 uL Elution Buffer, incubate at RT for 1 min, centrifuge for 1 min	Add 80 uL Elution Buffer, incubate at RT for 1 min, centrifuge for 1 min

After the samples were prepared, they were vortexed briefly and their concentrations were measured using the corresponding elution buffer (depending on which kit was used that day) as a blank with the Nanodrop One device manufactured by Thermo Scientific. Once the concentrations were measured, the samples were analyzed with a 1% agarose gel (1 g agarose in ~100 mL 0.5x TAE buffer with ~3 uL DNA Stain G). Samples were diluted with the following ratio: 3 uL 5x loading dye + 6 uL nuclease-free water + 1 uL sample DNA. Controls included CALR WT and CALR del52 plasmids. This gel was placed in an apparatus filled with 0.5x TAE buffer and ran at 100 V for ~45-60 minutes. Stained gels were visualized with the Vilber-Lourmat UV transilluminator. If the sample bands were of a similar size to either the CALR WT or del52 control plasmids, they were sent to MSM Ürün ve Hizmet (İstanbul, Türkiye) before going to MacroGen Europe BV (Amsterdam, Netherlands) for Sanger sequencing.

2.2.7) Sanger sequencing result analysis with Benchling

Once the sequencing results were obtained from MacroGen, the raw Sanger reads in .ab1 format were uploaded to Benchling and were compared to either CALR WT or CALR del52 (full sequences of each were uploaded separately) using the built-in MAFFT (multiple alignment using Fast Fourier transform) tool optimized for comparing nucleotide sequences, which takes the commonly shared nucleotides between the template and the individual samples. Sequences were screened for the presence of truncations or mutations (depending on the mutagenesis), and if confirmed, the plasmids were then used for transfecting HEK293 cells.

2.2.8) Calcium phosphate transfection of HEK293 cells

Two different reagents were prepared: calcium chloride and 2x HBS (Hanks' Balanced Salt) solution. For each transfection, a fresh 2 M calcium chloride solution was prepared (2.93 g in 10 mL Milli-Q water). The HBS solution was prepared as follows:

Reagent	Amount	Final Concentration
Sodium chloride (NaCl)	8.18 g / 500 mL	280 mM
Potassium chloride (KCl)	0.37 g / 500 mL	10 mM
Sodium phosphate dibasic (Na ₂ HPO ₄)	0.106 g / 500 mL	1.5 mM
Dextrose	1.08 g / 500 mL	12 mM
HEPES	5.96 g	50 mM (pH adjusted to 7.05)
Milli-Q Water	Fill to 500 mL	N/A

A 35 mm culture plate was seeded with approximately 300,000 cells with 2 mL complete DMEM media added to neutralize the trypsin. The next day, the following mixture was prepared to transfect the cells:

Reagent	Amount
2x HBS solution	125 uL
Transfected DNA	~1 ug
2 M CaCl ₂	Varies by DNA concentration
Filter-sterilized Milli-Q water	Fill to 125 uL (excluding HBS, such that it is diluted two-fold)

After vortexing the transfection mix, it was left to incubate at RT for 20 min. During this incubation, the old cell media was removed using a pre-sterilized glass pipette, the cells were then washed with ~2 mL PBS (phosphate buffer saline), the PBS was aspirated, and the cells were then

replenished with fresh media. Once the incubation was finished, the 250 uL transfection mix was added drip-wise over the plate such that the mix covered all the cells, which was then returned to the 37 degree incubator. After at least 24 hours, the transfected HEK cells were then collected for Western blot analysis.

2.2.9) Western blot protocol

Transfected cells were first washed twice with ice-cold PBS after the media was aspirated, then centrifuged at 1500 rpm for 2 minutes. After removing the PBS, the cells were resuspended with the following NP-40 lysis buffer (~1 mL) on ice:

Reagent	Amount
5 M NaCl	15 mL
10% Nonidet P-40 (NP-40)	50 mL
1 M Tris pH 8.0	25 mL
Milli-Q Water	Fill to 500 mL

The resuspended cells were then centrifuged at 1500 rpm at 4 degrees for 10 minutes. The supernatant was collected and moved to a -20 degree freezer until Western blots were ready to run. For the initial steps, two different gels were made for separation and stacking, respectively:

7 mL Separating Gel (12%)	
Water	2.34 mL
30% Bis-Acrylamide	2.8 mL
1.5 M Tris-HCL pH 8.8	1.75 mL
20% SDS (sodium dodecasulfate)	35 uL
10% APS (ammonium persulfate)	70 uL
TEMED	3 uL

3 mL Stacking Gel (5%)	
Water	2 mL
30% Bis-Acrylamide	500 uL
0.5 M Tris-Hcl pH 6.8	378 uL
20% SDS	15 uL
10% APS	30 uL
TEMED	3 uL

The samples were diluted with 4x Laemmli buffer and DTT (5 uL 4x Laemmli buffer, 5 uL DTT, 10 uL sample), denatured at 95 degrees Celsius for 5 minutes in a hot plate, then loaded into the Western blot gel tank filled with 1x running buffer. 4 uL of Opti-Marker protein ladder was used as a control. The samples were ran at 80 V until they passed the stacking gel (~20 minutes), and then 120 V until they reached the bottom of the separating gel (approximately one hour). The gel was then removed from the tank and kept in 1x transfer buffer (running buffer was collected for later reuse). The transfer cassette was assembled according to the Bio-Rad protocol and placed back in the tank with 1x transfer buffer and ran at 250 mA for 2 hours. The membrane was then washed and soaked with 1x TBS, stained with Ponceau S solution, which was then washed three times with TBST (TBS plus Tween 20). Blocking buffer (30 mL TBST plus 3 g dry milk powder) was added to the membrane, which was incubated for 30 minutes. The membrane was washed with TBST for 6 minutes twice, then washed with TBS three times for 5 minutes each. The primary antibody (CALR) was prepared in a 1:3,000 solution (12 mL TBS, 4 uL antibody, and 12 uL 10% sodium azide), and the membrane was incubated overnight at 4 degrees Celsius soaked inside a 50 mL Falcon tube on an incubator-shaker. The next day, the membrane was washed three times with TBST three times for 7 minutes each. Next, the

secondary antibody (anti-rabbit) was prepared in a 1:100,000 solution (0.5 uL anti-rabbit antibody in 5 mL TBST), with the membrane incubated at RT for one hour. The membrane was washed with TBST three times for 7 minutes each, then with TBS for 5 minutes. 800 uL Glossy ECL HRP (horseradish peroxidase) was then added to the membrane, which was incubated in the dark for two minutes. Bands were then visualized using the Amersham Imager 600 (Amersham, United Kingdom). The membrane was then washed with 1x stripping buffer twice for 10 minutes each time. Blocking buffer was added as before in order to add the 1:100,000 anti-beta actin antibody solution (4 mL TBS + 0.4 uL antibody + 10% sodium azide), and the steps were repeated all the way up to visualization with the Amersham Imager 600.

2.2.10) Obtaining PDB files for CALR and visualization with PyMOL

PyMOL was used to visualize PDB files obtained from LocalColabFold [113], where single residues were mutated to different types of residues corresponding to *in silico* predictions, as well as to record distances between atoms of select residues (in angstroms). LocalColabFold works by using the sequence of the protein of interest as an input, then uses MMSeq2 homology search to find the closest protein in terms of sequence (in this case, WT CALR) and then generates five PDB files ranked by pLDDT (predicted local distance difference test). The highest-ranked PDB file according to pLDDT was selected for PyMOL visualizations and measurement heatmaps, which were done with the seaborn package in Python as before.

2.2.11) Predicting post-translational modification on select residues on CALR del52

When predicting for different types of post-translational modifications (PTM), there were two different tools used: web-server MuSiteDeep [115] for predicting levels of phosphorylation and O-linked glycosylation by residue, and DeepPhos [116] on GitHub as a comparison for MuSiteDeep phosphorylation prediction data. MuSiteDeep uses a convolutional neural network (CNN) that also employs UniProt and SwissProt database proteins to continuously learn and parse out patterns from protein sequences to predict several types of PTMs. Although only the protein sequence was the initial input, the corresponding 3-D structure was also factored in to estimate probabilities of PTMs occurring at certain residues by employing structures found in the RCSB database. Similar to MuSiteDeep, DeepPhos also utilizes a CNN model, but is specifically tailored towards identifying potential sites for different kinase families and residue types (serine or threonine, no tyrosines were present in the mutant CALR C-terminus). Based on select residues that showed trends in either direction for these modifications, further predictions were made based on the mutations of these residues to alanines to see if the presence of these post-translational modifications hindered or facilitated predicted secretion. Heatmaps were then used to compare effects from the mutations of these residues as well as combinations of these mutations to see if there was an accumulative effect on predicted secretion or not.

CHAPTER 3

RESULTS

3.1) DeepLoc sorting signal results

The first analysis was analyzing the CALR del52, ins5, and WT sequences through the DeepLoc search engine to see the importance of each amino acid in the sequence in terms of general sorting signals. In the case of CALR WT, the KDEL residues at the very end of the C-terminus did appear on the graph, as expected, to be highly important. Additionally, the first fifteen residues constituting the N-terminal signal peptide that directs the protein towards the ER [78] also had a high score (see Figures 7a-7c). For both CALR mutants, the predicted sorting signal importance for the N-terminal almost doubled compared to WT, while some of the proposed motifs appeared in the search results.

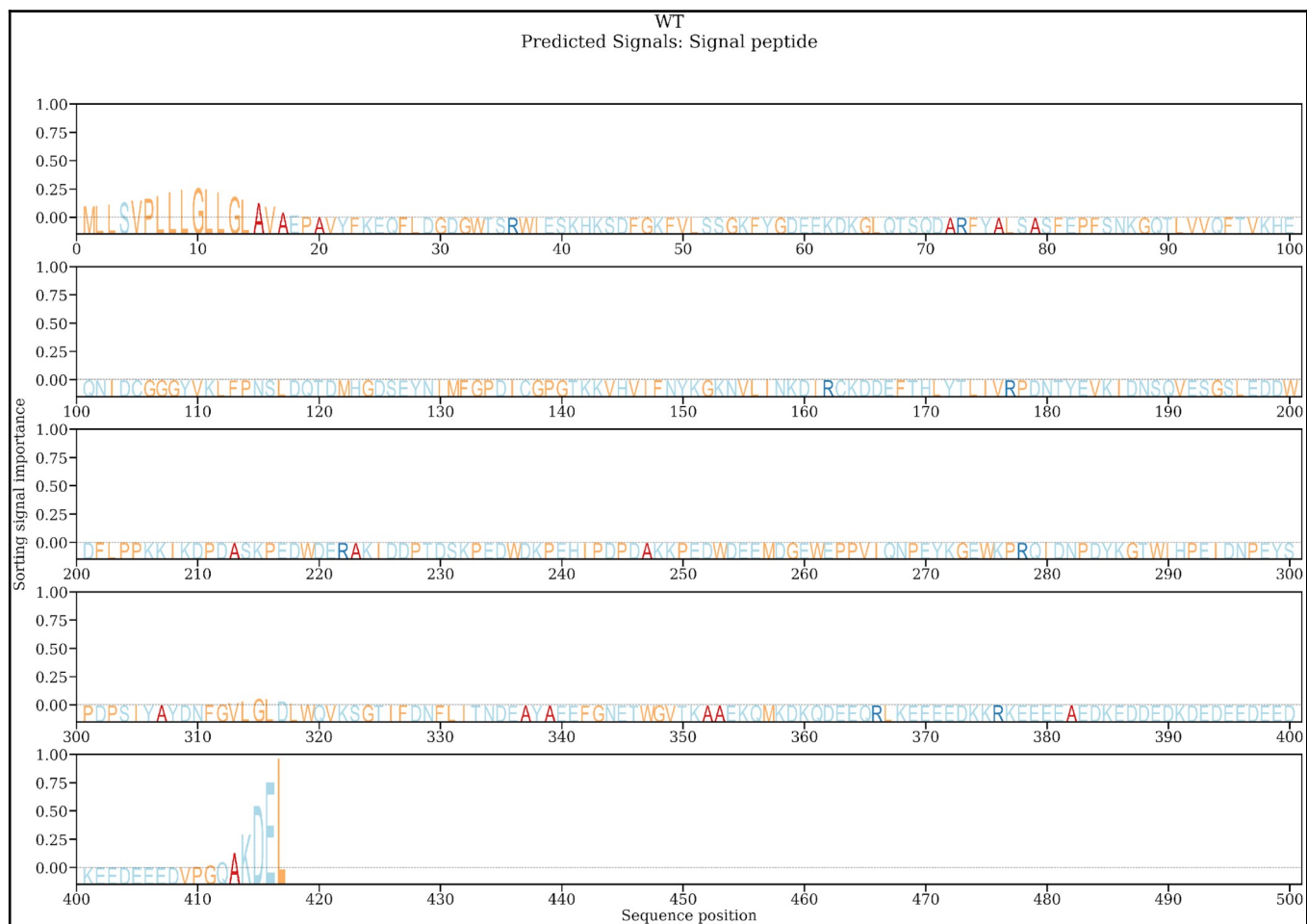


Figure 7a. Predicted sorting signal importance for each amino acid in the CALR WT sequence, generated by DeepLoc.

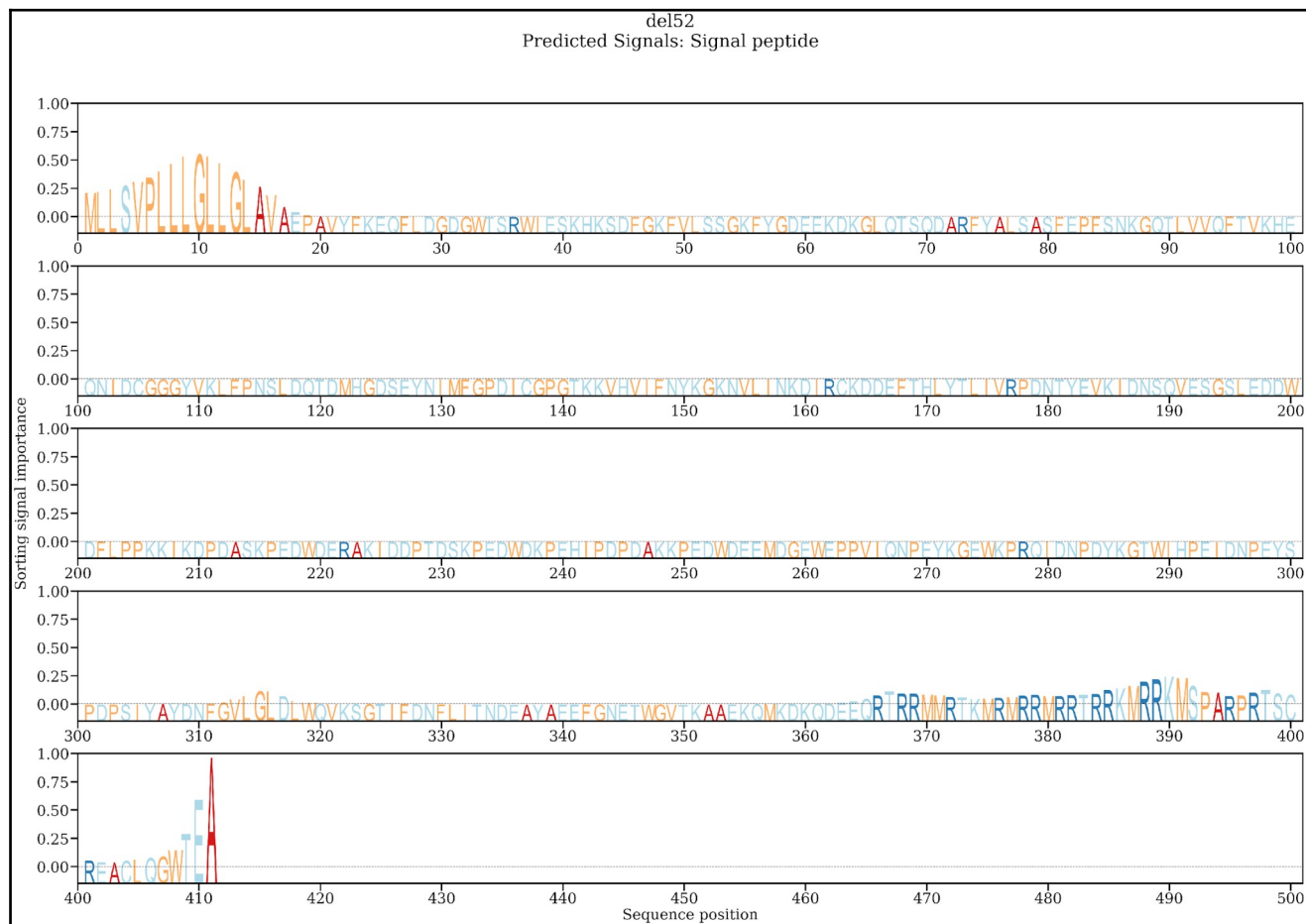


Figure 7b. Predicted sorting signal importance for each amino acid in the CALR del52 sequence, generated by DeepLoc.

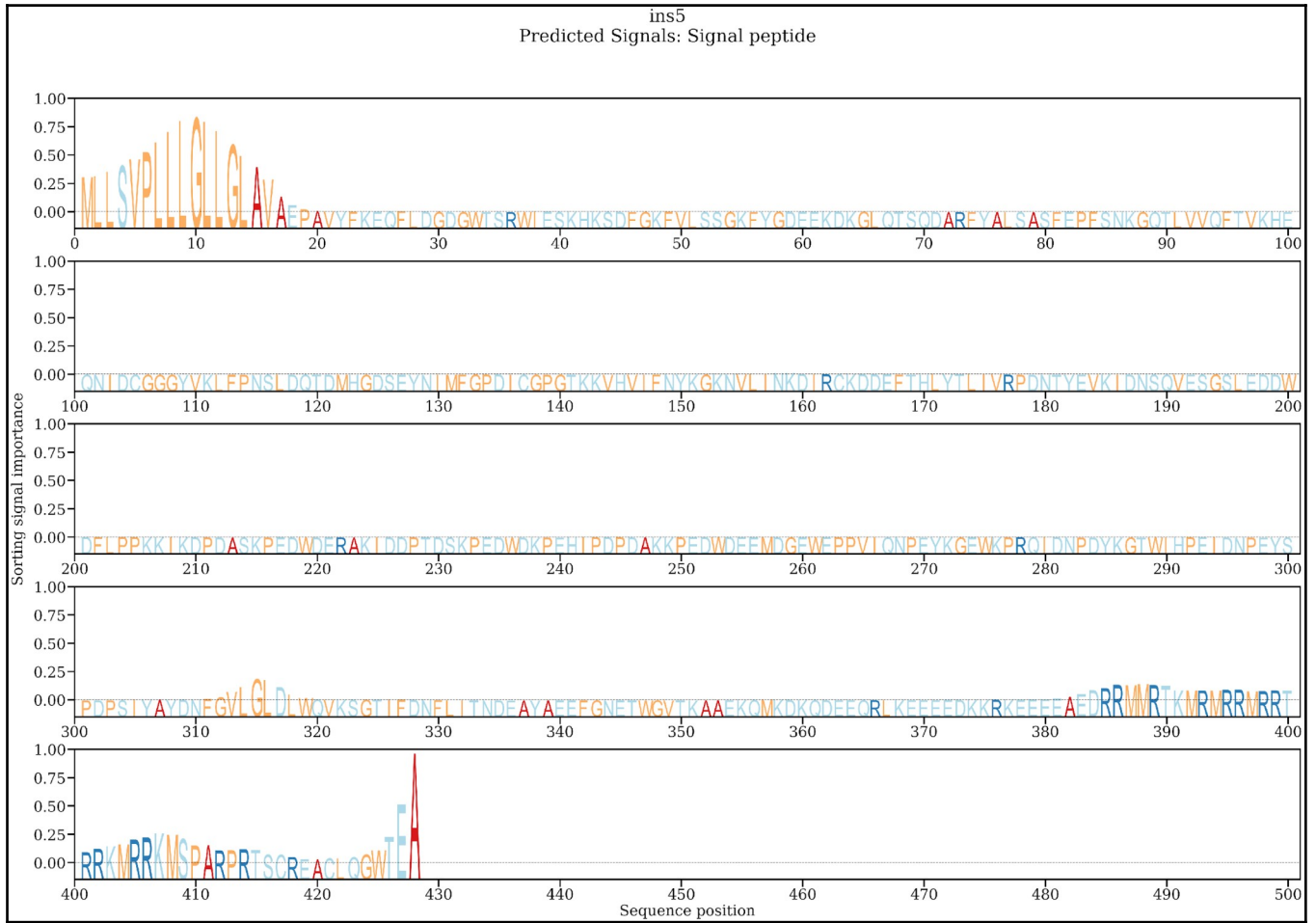


Figure 7c. Predicted sorting signal importance for each amino acid in the CALR ins5 sequence, generated by DeepLoc.

3.2) ELM search results for CALR del52

The proposed secretion motifs seen in Table 3c also had some predicted sorting signal importance, albeit lower than either the mutant or WT C-termini, or the common N-terminus. Between del52 and ins5, the arginine-rich portion of the mutant C-terminus had a higher predicted importance score for del52. However, the last four residues for either mutant (-WTEA) also appeared with a significant score, but this had no immediate answer as these residues have no known association with the changes associated with these frameshift mutations. To help account for this result, and to check for any organelles that may be correlated with these motifs, our next analysis was using the ELM (Eukaryotic Linear Motifs) search engine (elm.eu.org) to parse out for possible motifs and their associated localizations. Furthermore, if the motif also formed part of a protease cleavage site, this type of information would also appear in the search result as proteases have defined cellular localizations. For many of our proposed motifs, they did appear in the search result, with the general localization being either the extracellular matrix, Golgi apparatus, or cell membrane (see Table 4a). Intriguingly, the last seven residues -LQGWTEA also appear as a predicted docking site for the phosphatase PP2A inside the nucleus, thus accounting for the DeepLoc results seen in Figures 7b and 7c. Indeed, many of the motifs, including most of them combined (see Table 4b) are also affiliated with nuclear localization, which is consistent with prior findings on the over-representation of lysines and arginines in nuclear localization signals [117]. Mutant CALR has been found to be disproportionately found in the nucleus as well as the extracellular space compared to WT [107]. Although there was still some association with the ER, the results showed that the motifs were more likely to be involved in ER-Golgi trafficking rather than retention in the ER, which is also consistent with our current understanding of the secretion mechanism of mutant CALR. No published literature was factored into the calculation of

likelihood for these motifs, they were based on the location of the motif along the sequence and location of potential proteases in the cell. However, the same type of search was performed for WT CALR as a control, and the KDEL motif did appear in the results, but as a true positive, meaning that there was explicit experimental confirmation (see Table 4c). This is also why the table is differently colored in dark orange unlike the pale cream tint seen for the arginine-rich motifs.

Table 4a. Predicted motifs from CALR del52 and their predicted cellular localization. Predictions done using the entire del52 sequence in the search engine from Eukaryotic Linear Motifs for Functional Sites in Proteins [108]. There are three filters used by the ELM motif search engine: globular domain, 3-D structure, and context in terms of position on the amino acid sequence. Abbreviations on the left side: CLV: cleavage site, NRD: N-arginine dibasic convertase, PCSK: proprotein convertase 7, KEX: kexin, FUR: furin.

CLV_NRD_NRD_1	RRM RRM RRT RRK RRK	368-370 [A] 378-380 [A] 381-383 [A] 384-386 [A] 388-390 [A]	- - - - -	N-Arg dibasic convertase (NRD/Nardilysin) cleavage site (X- -R- K or R- -R-X).	extracellular, Golgi apparatus, cell surface
CLV_PCSK_FUR_1	RTRRM RMRRM RMRRT RTRRK	366-370 [A] 376-380 [A] 379-383 [A] 382-386 [A]	- - - -	Furin (PACE) cleavage site (R-X-[RK]-R- - X).	extracellular, Golgi apparatus, Golgi membrane
CLV_PCSK_KEX2_1	RRM RRM RRT RRK RRK	368-370 [A] 378-380 [A] 381-383 [A] 384-386 [A] 388-390 [A]	- - - - -	Yeast kexin 2 cleavage site (K-R- - X or R-R- -X).	extracellular, Golgi apparatus
CLV_PCSK_PC7_1	RRKMRRK	384-390 [A]	-	Proprotein convertase 7 (PC7, PCSK7) cleavage site (R-X-X-X-[RK]-R- - X).	extracellular, Golgi apparatus, Golgi membrane

Table 4b. Combination of predicted motifs in the del52 sequence suggested to play a role in either nuclear localization or secretion from the ER. Predictions also generated by ELM. Abbreviations on the left side: DOC: docking site, NLS: nuclear localization signal, and TRG: targeting motif.

DOC_PP2A_B56_1	LQGWTEA	405-411 [A]	-	Docking site required for the regulatory subunit B56 of PP2A for protein dephosphorylation.	nucleus, kinetochore, chromosome, centromeric region, cytosol
TRG_ER_diArg_1	RTRR RRM MRMR RMRR MRR RRMR RMRR MRR RRTR RTRR MRR	366-369 [A] 368-370 [A] 375-378 [A] 376-379 [A] 377-379 [A] 378-381 [A] 379-382 [A] 380-382 [A] 381-384 [A] 382-385 [A] 387-389 [A]	- - - - - - - - - - -	The di-Arg ER retention motif is defined by two consecutive arginine residues (RR) or with a single residue insertion (RXR). The motif is completed by an adjacent hydrophobic/arginine residue which may be on either side of the Arg pair.	cytosol, endoplasmic reticulum membrane, integral protein, ER-Golgi transport vesicle membrane, endoplasmic reticulum membrane, Golgi-ER transport vesicle membrane, rough endoplasmic reticulum, endoplasmic reticulum cisterna
TRG_NLS_Bipartite_1	RRMMRTKMRMRMRTRRKMR	368-389 [A]	-	Bipartite variant of the classical basically charged NLS.	nucleus, Nuclear pore, NLS-dependent protein nuclear import complex
TRG_NLS_MonoExtC_3	RRKMRR	384-389 [A]	-	Monopartite variant of the classical basically charged NLS. C-extended version.	nucleus, Nuclear pore, NLS-dependent protein nuclear import complex
TRG_NLS_MonoExtN_4	RTRRKMR RRKMRR	382-389 [A] 384-389 [A]	- -	Monopartite variant of the classical basically charged NLS. N-extended version.	nucleus, Nuclear pore, NLS-dependent protein nuclear import complex

Table 4c. CALR WT anchor motif KDEL shown in ELM results. This motif is highlighted in orange because of its extensive documentation, hence the “true positive” label.

Elm Name	Instances (Matched Sequence)	Positions	Logic	Elm Description	Cell Compartment	Pattern
TRG_ER_KDEL_1	KDEL	414-417	true positive	Golgi-to-ER retrieving signal found at the C-terminus of many ER soluble proteins. It interacts with the KDEL receptor which in turns interacts with components of the coatomer (COP I).	endoplasmic reticulum lumen, Golgi lumen, rough endoplasmic reticulum, endoplasmic reticulum cisterna, Golgi apparatus	[KRHQSAP][DENQT]ELS

3.3) Individual motif mutageneses for CALR WT and del52 in DeepLoc and MULocDeep

Following these results, the proposed motifs were mutated to alanines to see if they made an impact on predicted secretion. In CALR WT, negatively charged motifs were mutated to residues similar to what is in the mutant C-terminus to see if there was an increase in secretion. As seen in Table 5a-b, three residues at a time were mutated to motifs resembling those seen in mutant CALR. The exception is the KRK mutated to alanines, which serves to act as a control to see if the loss of positively charged residues contributes to a loss in predicted secretion.

Protein sequences were processed using DeepLoc and MULocDeep to determine the effect on predicted localization either in the extracellular matrix for MULocDeep or in the “secreted” category in the case of DeepLoc. For CALR del52, the motif which had the greatest effect on secretion was RRM1 when mutated to alanines. RRM1 also had one of the largest drops in secretion but it was smaller according to DeepLoc: MULocDeep predicted a 35% drop in secretion compared to less than 10% for DeepLoc. Some motifs resulted in a modest increase in secretion: RTKMR-AAAAA and RRM2-AAA

had higher levels of secretion than del52 according to MULocDeep, but this was not the case with DeepLoc: instead, there was a modest decrease in secretion. For CALR WT, individual mutations generally increased predicted secretion but none of them approached the level of del52 (Figure 8c).

Table 5a-5b. List of performed mutations for CALR del52 (a) and WT (b). Figure shows the C-terminal protein sequence for each variant after mutagenesis is successful. For the CALR WT, the motifs are intended to recreate the mutant C-terminus, and for the CALR del52, the goal is to neutralize the positively charged motifs.

mutagenesis_ID	sequence
RRK1_AAA	RTRRMMRTKMRMRMRRTTRRKMAAAMSPARPRTSCREACLQGWTEA
RRK2_AAA	RTRRMMRTKMRMRMRRTAAAMRRKMSPARPRTSCREACLQGWTEA
RRT_AAA	RTRRMMRTKMRMRMAAARRKMRRKMSPARPRTSCREACLQGWTEA
RRM_AAA	RTRRMMRTKMRMAAARRTRRKMRKMSPARPRTSCREACLQGWTEA
MRTKMR_AAAAAA	RTRRMAAAAAMRRMRRTTRRKMRKMSPARPRTSCREACLQGWTEA
RRM2_AAA	RRTAAAMRTKMRMRMRRTTRRKMRKMSPARPRTSCREACLQGWTEA
Del52	RTRRMMRTKMRMRMRRTTRRKMRKMSPARPRTSCREACLQGWTEA

mutagenesis_ID	Sequence
CALR_WT	LKEEEEDKKRKEEEEAEDKEDDEDKDEDEEDEDKEEDEEEDVPGQAKDEL
KRK_AAA	LKEEEEDKAAAEDEEAEDKEDDEDKDEDEEDEDKEEDEEEDVPGQAKDEL
DEE_RRK	LKEEEEDKKRKEEEEAEDKEDDEDKDEDEERRKDEEDEEEDVPGQAKDEL
DKD_RRK	LKEEEEDKKRKEEEEAEDKEDDERRKDEDEEDEDKEEDEEEDVPGQAKDEL
EDD_RRT	LKEEEEDKKRKEEEEAEDKRRTEDKDEDEEDEDKEEDEEEDVPGQAKDEL
EDK_RRM	LKEEEEDKKRKEEEEAEDKEDDRRMDEDEEDEDKEEDEEEDVPGQAKDEL
EEEEA_RTKMR	LKEEEEDKKRKRTKMRDEKEDDEDKDEDEEDEDKEEDEEEDVPGQAKDEL

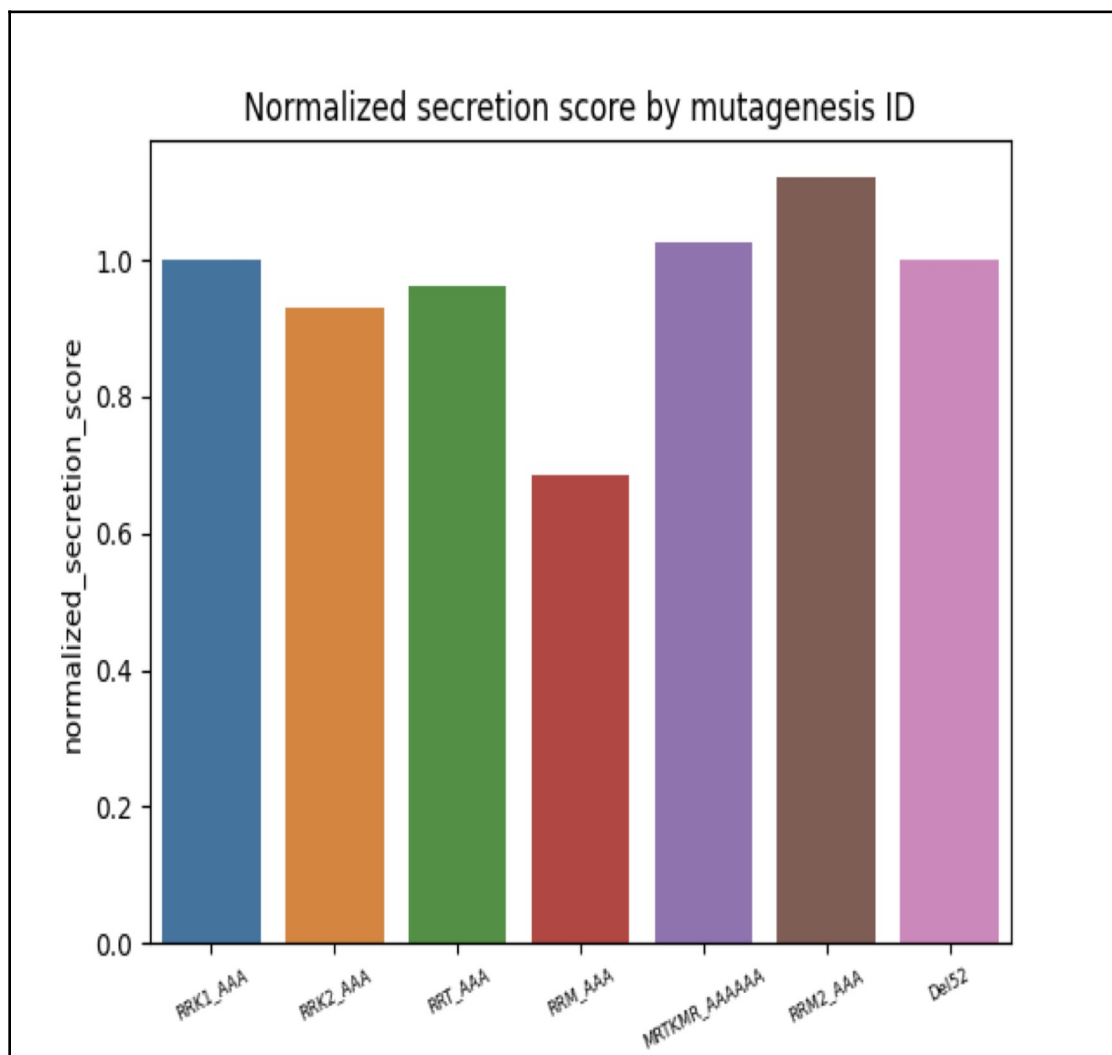


Figure 8a. Predicted secretion scores for planned CALR del52 mutations according to MULocDeep, normalized to del52.

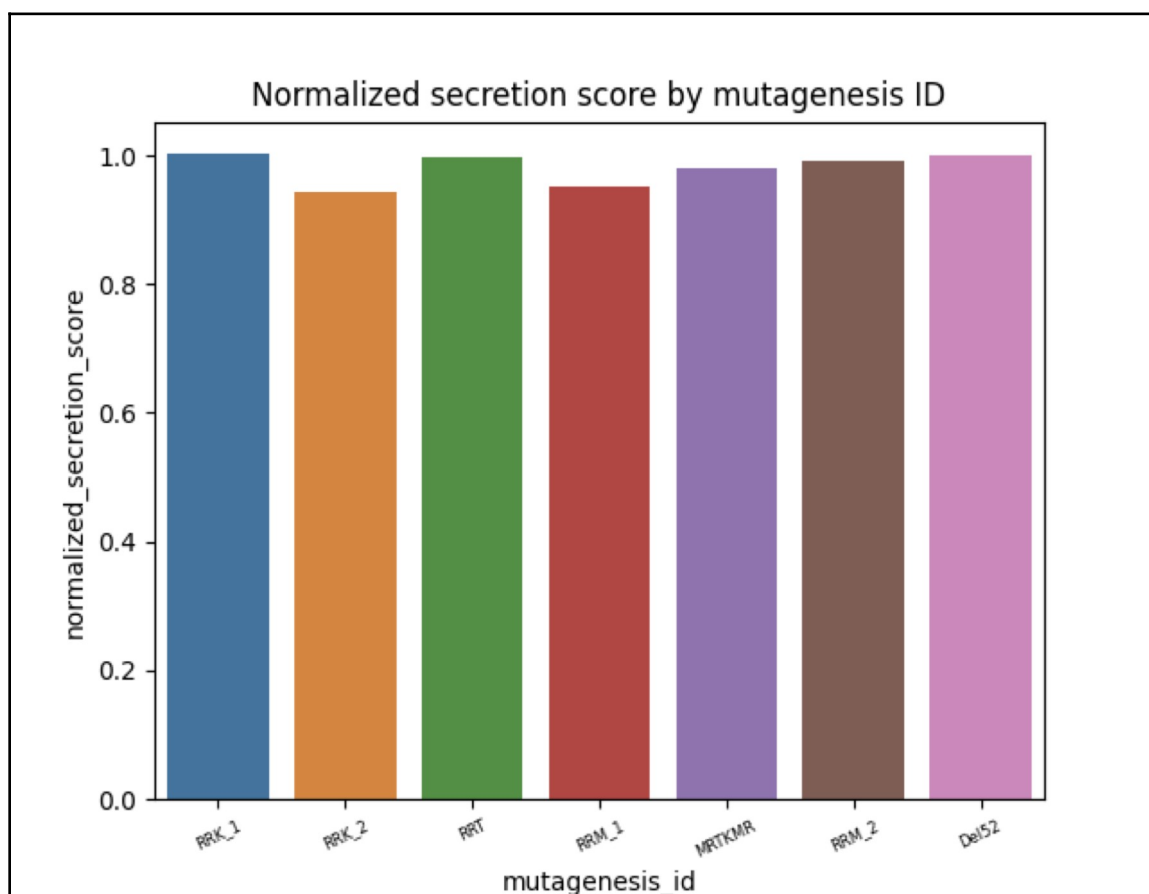


Figure 8b. Predicted secretion scores for planned CALR del52 mutations according to DeepLoc, normalized to del52.

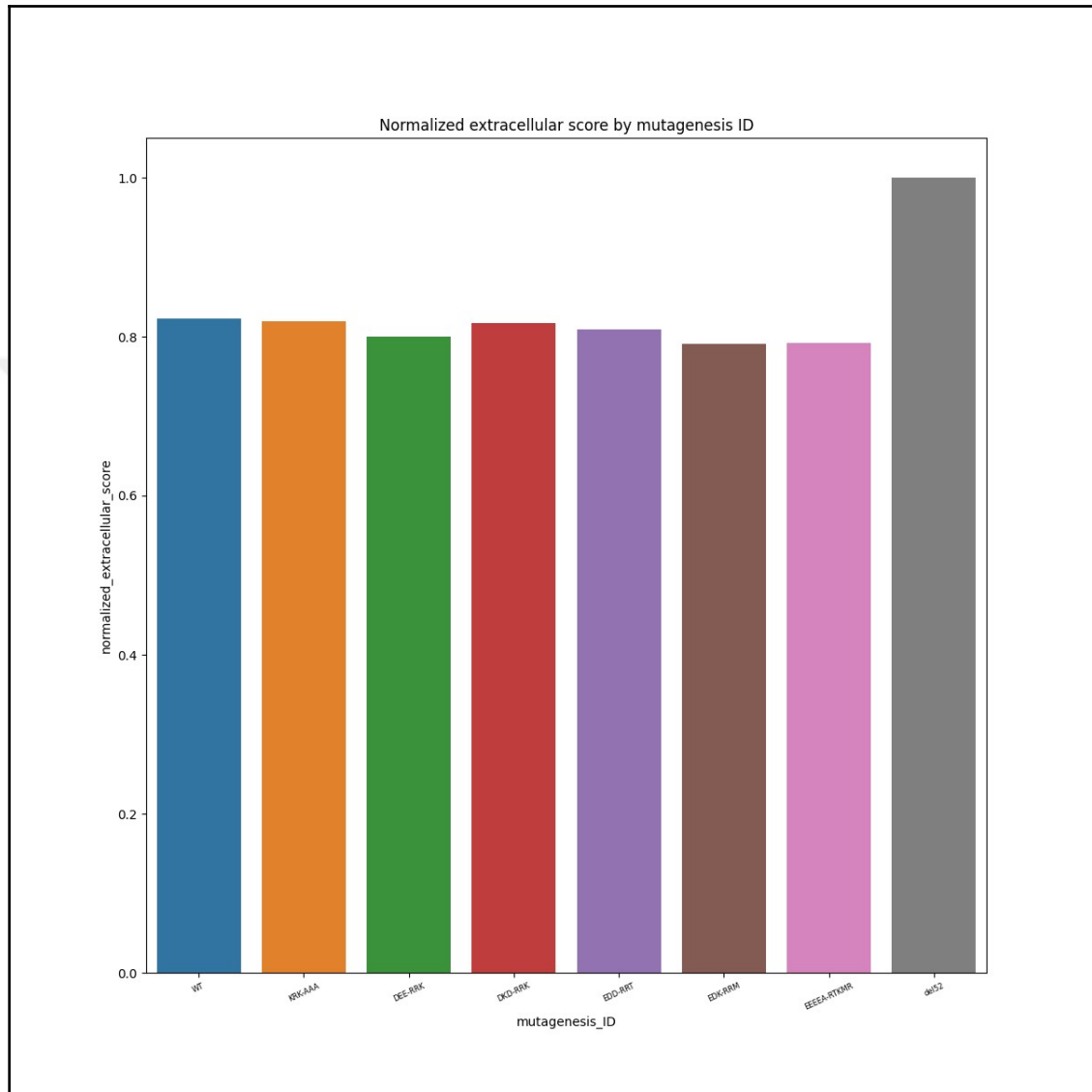


Figure 8c. Predicted extracellular scores for planned CALR WT mutations according to DeepLoc, with scores normalized to CALR del52.

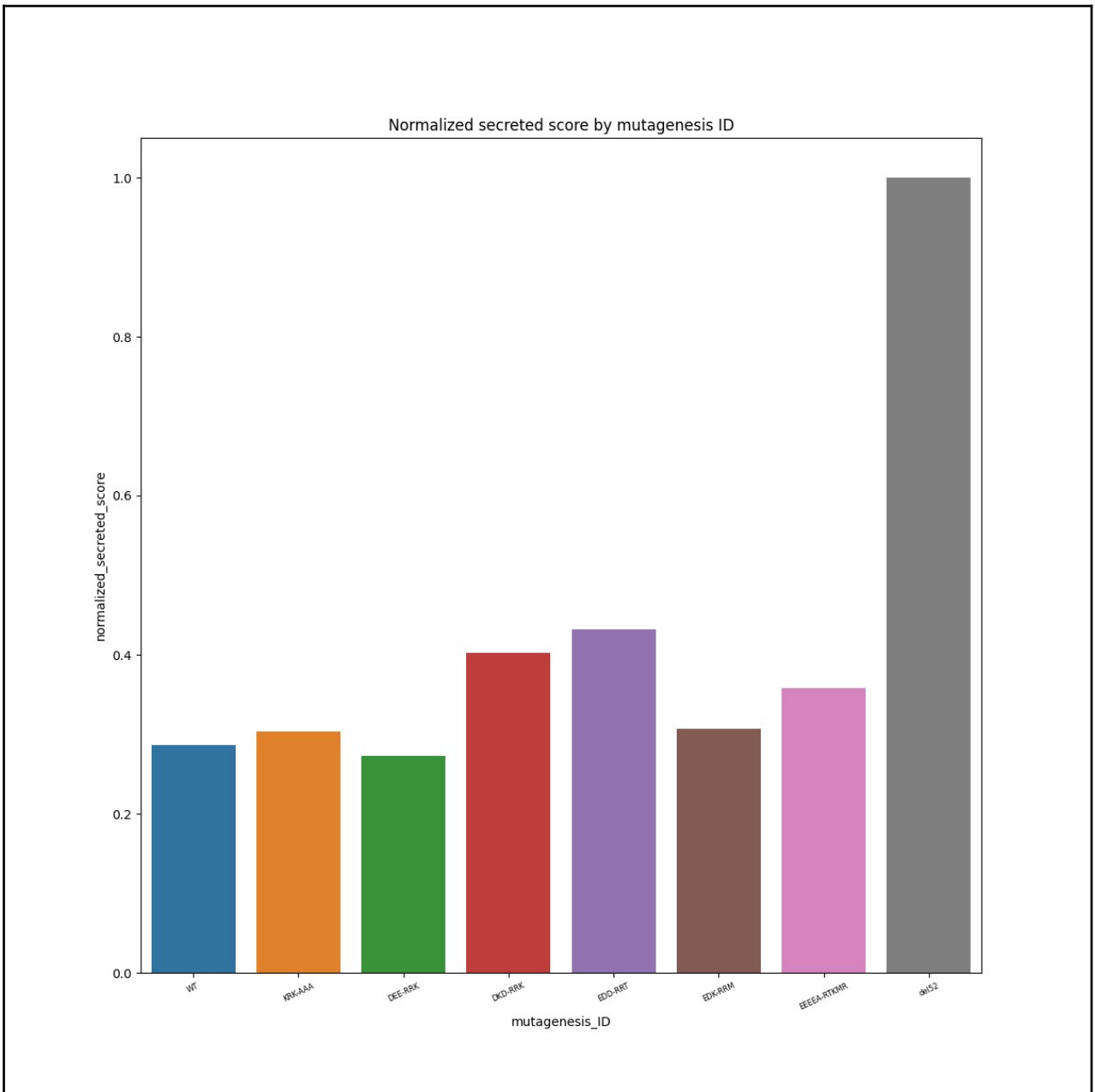


Figure 8d. Predicted extracellular scores for planned CALR WT mutations according to MULocDeep, with scores normalized to CALR del52.

3.4) Combinations of del52 and WT mutageneses in DeepLoc and MULocDeep

Going beyond individual mutations, combinations of mutations were also tested in both DeepLoc and MULocDeep to see if there was a cumulative effect in terms of decreasing secretion for del52 or increasing secretion for wild type (see Table 6 for more details). As more motifs were mutated to alanines in del52, the predicted secretion decreased significantly in both tools, although the decrease was more modest according in DeepLoc (Figures 9a and 9b). For CALR WT, the cumulative presence of the secretion motifs increased the score, but the combination of all these motifs was still lower than del52 (Figures 9c and 9d), suggesting that the KDEL retention signal may act as a counterbalance to these motifs [107]. However, for both of these predictions, DeepLoc is more conservative in its estimates: the maximum drop in secretion is approximately 20% for all CALR del52 motifs mutated to alanines (Figure 9b), while MULocDeep estimated a ~90% drop in secretion. Interestingly, for the CALR del52 mutations, there was a steeper drop when RRK1, RRK2, RRT, and RRM2 were mutated to alanines compared to RTKMR and RRM2 according to MULocDeep (see Figure 9a). However, this was not the case for DeepLoc, which estimated a similar drop in secretion when motifs were sequentially mutated from either the C-terminal or the N-terminal direction (Figure 9b).

In addition to these *in silico* predictions, preliminary data shows in both Western blots that the loss of these motifs results in less secretion though greater retention inside the cell, as seen by higher intensity bands inside the cell lysate (see Figure 9c). In combination with what was observed in the DeepLoc and MULocDeep predictions, the means of loss of these motifs is likely not important, as mutating at least some of the motifs (before M377, see Table 2 for full del52 sequence) to methionines had a similar result in stopping secretion as truncating the motifs. Mutating all residues past M377 to arginines resulted in increased secretion somewhat based on the intensity of the band, but it was still

stronger than the del52 band, indicating that some of it was still retained inside the cell. There is also evidence that accumulated mutations or larger truncations are needed to significantly halt CALR del52 secretion (compare with Figures 9a and 9b). For CALR WT, the accumulation of mutations towards the N-terminus had a greater impact on increasing secretion according to both DeepLoc and MULocDeep, although this effect was predicted to be more modest according to the former (compare Figures 9d and 9e). However, the addition of several motifs to the WT sequence was not sufficient to achieve predicted secretion levels identical to del52: approximately 50% of del52 secretion according to MULocDeep and 90% based on DeepLoc predictions.

Table 6. Diagrams showing different combinations of mutations for CALR WT (top) with the del52 C-terminus added as a reference for normalization, and CALR del52 mutations (bottom). KRK mutated to alanines as a control, but retained in other mutagenesis combinations to better resemble the mutant C-terminus. Mutated residues highlighted in dark blue.

del52	RTRRMMRTKMRMRMRRTTRRKMRK MSPARPRTSCREACLQGWTEA
RRK1	R TTRRMMRTKMRMRMRRTTRRK M AAA MSPARPRTSCREACLQGWTEA
RRK1_+_RRK2	RTRRMMRTKMRMRMRRT AAAMAAA MSPARPRTSCREACLQGWTEA
RRK1_+_RRK2_+_RRT	RTRRMMRTKMRMRMR AAAAAAMAAA MSPARPRTSCREACLQGWTEA
RRK1_+_RRK2_+_RRT_+_RRM	RTRRMMRTKMRM AAAAAAAAAMAAA MSPARPRTSCREACLQGWTEA
RRK1_+_RRK2_+_RRT_+_RRM_+_MRTKMR	RTRRM AAAAAAAAMAAAAAAAAAMAAA MSPARPRTSCREACLQGWTEA
all	RT AAAAAAAAAMAAAAAAAAAMAAA MSPARPRTSCREACLQGWTEA
RRM2_+_MRTKMR_+_RRM1_+_RRT_+_RRK2	RT AAAAAAAAAMAAAAAAAAAMRRK MSPARPRTSCREACLQGWTEA
RRM2_+_MRTKMR_+_RRM1_+_RRT	RT AAAAAAAAAMAAAAA RRKMRRK MSPARPRTSCREACLQGWTEA
RRM2_+_MRTKMR_+_RRM1	RT AAAAAAAAAMAAA RTRRKMRK MSPARPRTSCREACLQGWTEA
RRM2_+_MRTKMR	RT AAAAAAAAAMRRMRRTTRRKMRK MSPARPRTSCREACLQGWTEA
RRM2	RT AAAM RTKMRMRMRRTTRRKMRK MSPARPRTSCREACLQGWTEA

mutagenesis_ID	sequence
WT	L KEEEEDKKRKEEEE AED K EDDEDKDEDEEDEDKEEDEDVPGQAKDEL
KRK-AAA	L KEEEEDK AAAA EEEEE AED K EDDEDKDEDEEDEDKEEDEDVPGQAKDEL
DEE-RRK	L KEEEEDKKRKEEEE AED K EDDEDKDER RRK DEEDKEEDEDVPGQAKDEL
DEE-RRK DKD-RRK	L KEEEEDKKRKEEEE AED K EDDER RRKERRK DEEDKEEDEDVPGQAKDEL
EDD-RRT DKD-RRK DEE-RRK KRK	L KEEEEDKKRKEEEE AED KRRTERRKERRK DEEDKEEDEDVPGQAKDEL
EDD-RRT RRK DEE-RRK KRK EDK-RRM	L KEEER RRM KRKEEEE AED KRRTERRKERRK DEEDKEEDEDVPGQAKDEL
EDD-RRT DKD-RRK DEE-RRK KRK EDK-RRM EEEE-RTKMR	L KEEER RRM KRK RTKMRED KRRTERRKERRK DEEDKEEDEDVPGQAKDEL
EDD-RRT DKD-RRK KRK EDK-RRM EEEE-RTKMR	L KEEER RRM KRK RTKMRED KRRTERRK DEEDEDKEEDEDVPGQAKDEL
EDD-RRT KRK EDK-RRM EEEE-RTKMR	L KEEER RRM KRK RTKMRED KRRTERRK EDKDEEDEDKEEDEDVPGQAKDEL
KRK EDK-RRM EEEE-RTKMR	L KEEER RRM KRK RTKMRED K EDDEDKDEDEEDEDKEEDEDVPGQAKDEL
KRK EDK-RRM	L KEEER RRM KRKEEEE AED K EDDEDKDEDEEDEDKEEDEDVPGQAKDEL
del52	T RRMMRTKMRMRMRRTTRRKMRK MSPARPRTSCREACLQGWTEA

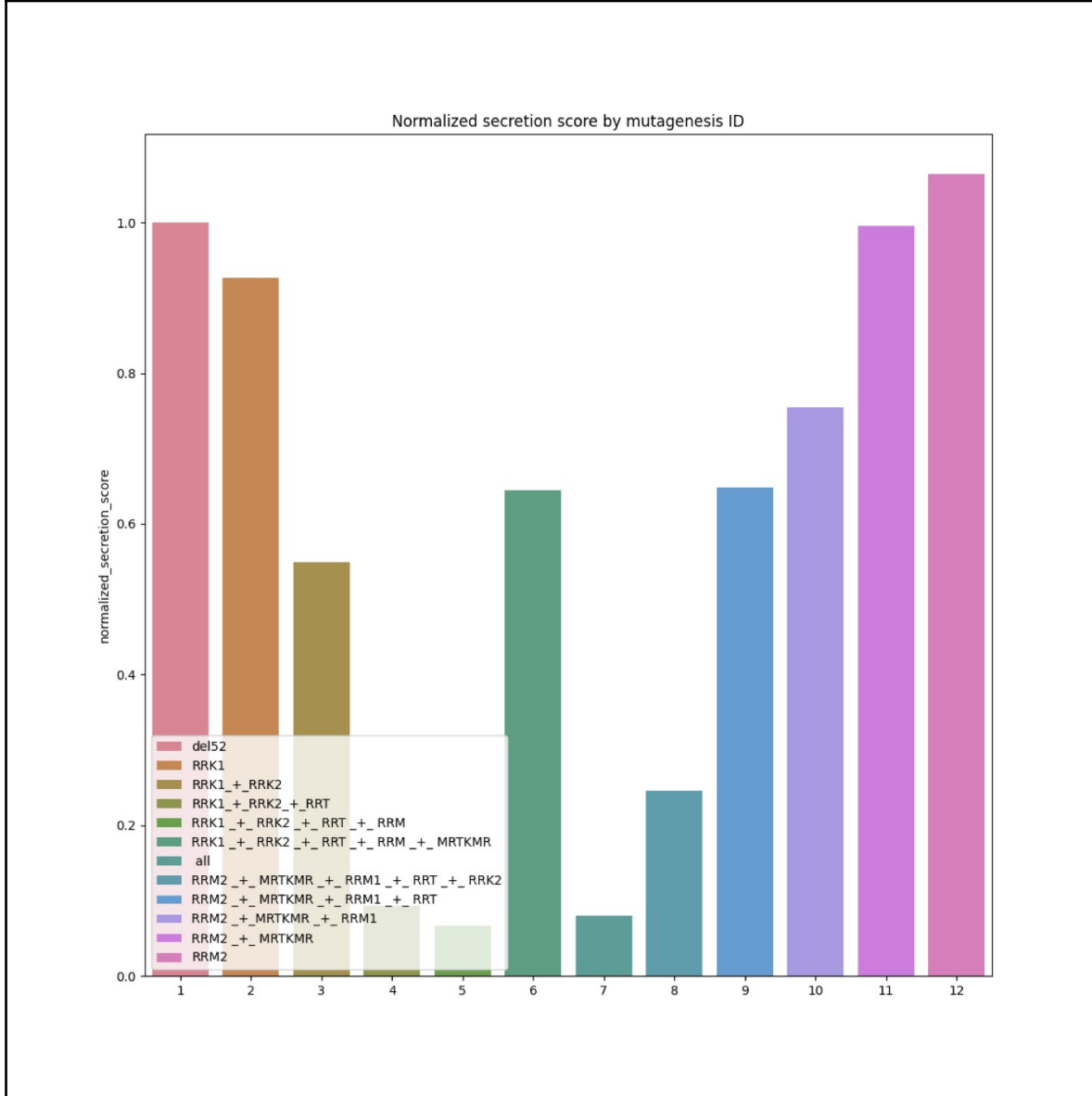


Figure 9a. Predicted secretion scores of different CALR del52 mutagenesis combinations normalized to CALR del52 ($x=1$) according to MULocDeep.

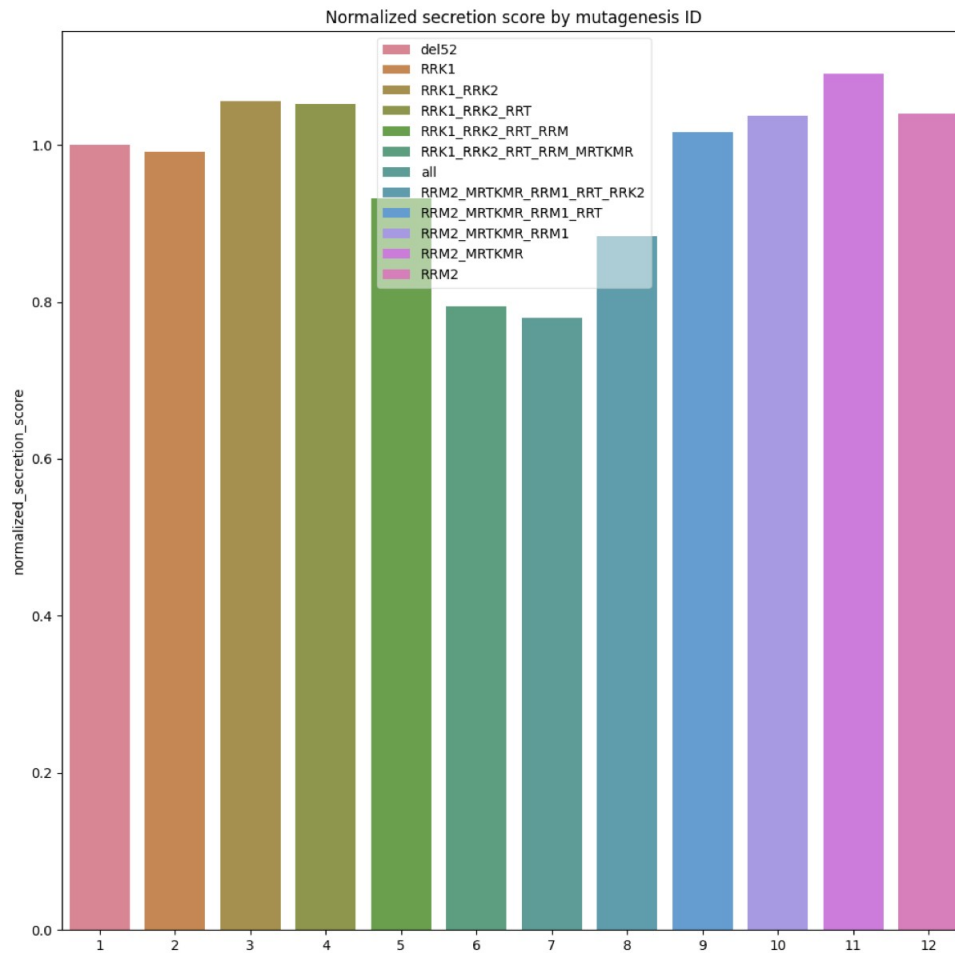


Figure 9b. Predicted secretion scores of different CALR del52 mutagenesis combinations normalized to CALR del52 ($x=1$) according to DeepLoc.

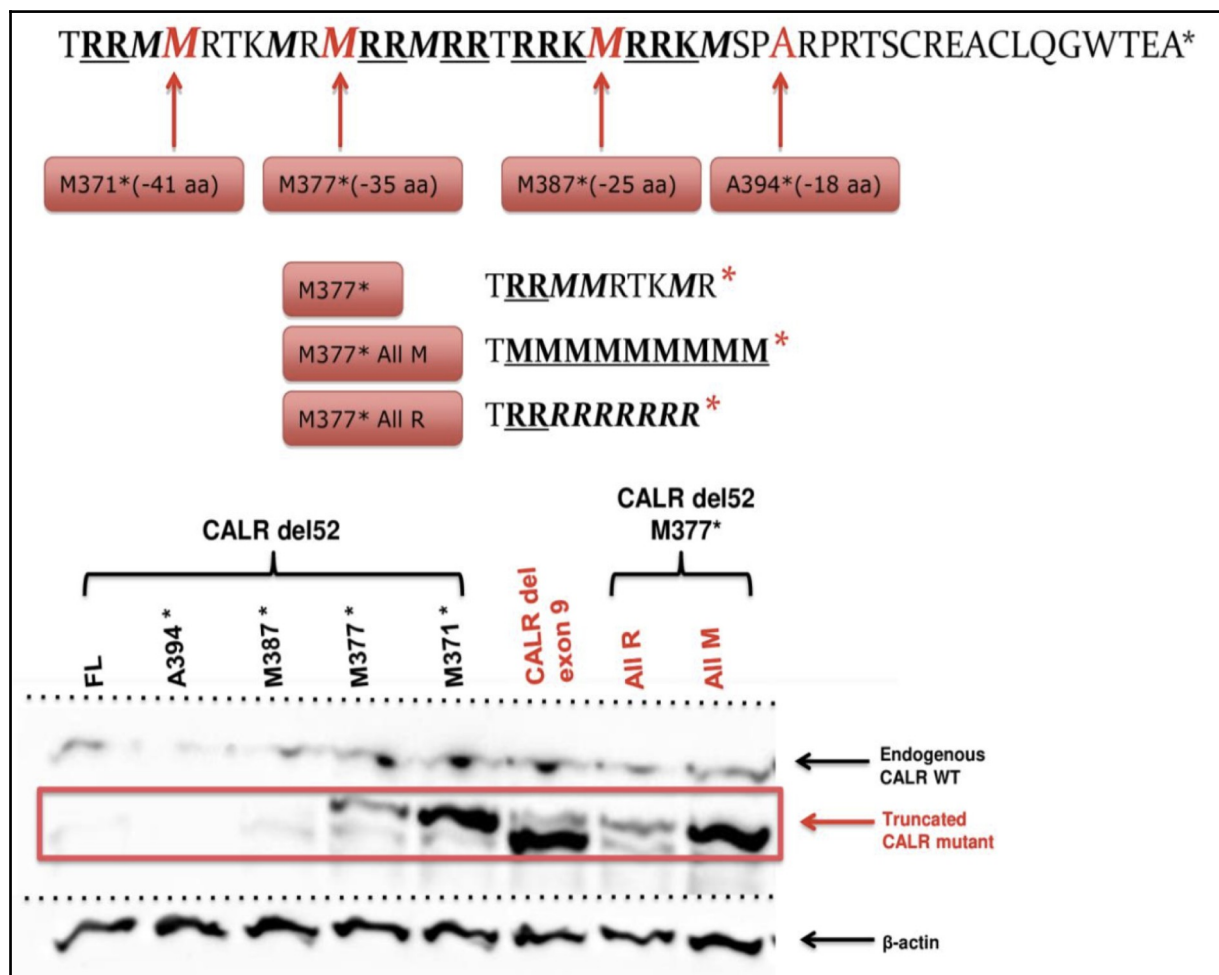


Figure 9c. Preliminary Western blot data comparing amounts of secreted mutant CALR via intensity of bands. CALR del exon 9 refers to truncation of all residues seen at the top of the diagram, while “All R” refers to mutating all residues past M377 to arginines, “All M” refers to mutating all residues past M377 to methionines. M377* refers to the introduction of a stop codon. Beta-actin used as a control, with truncations occurring at select residues marked with an asterisk.

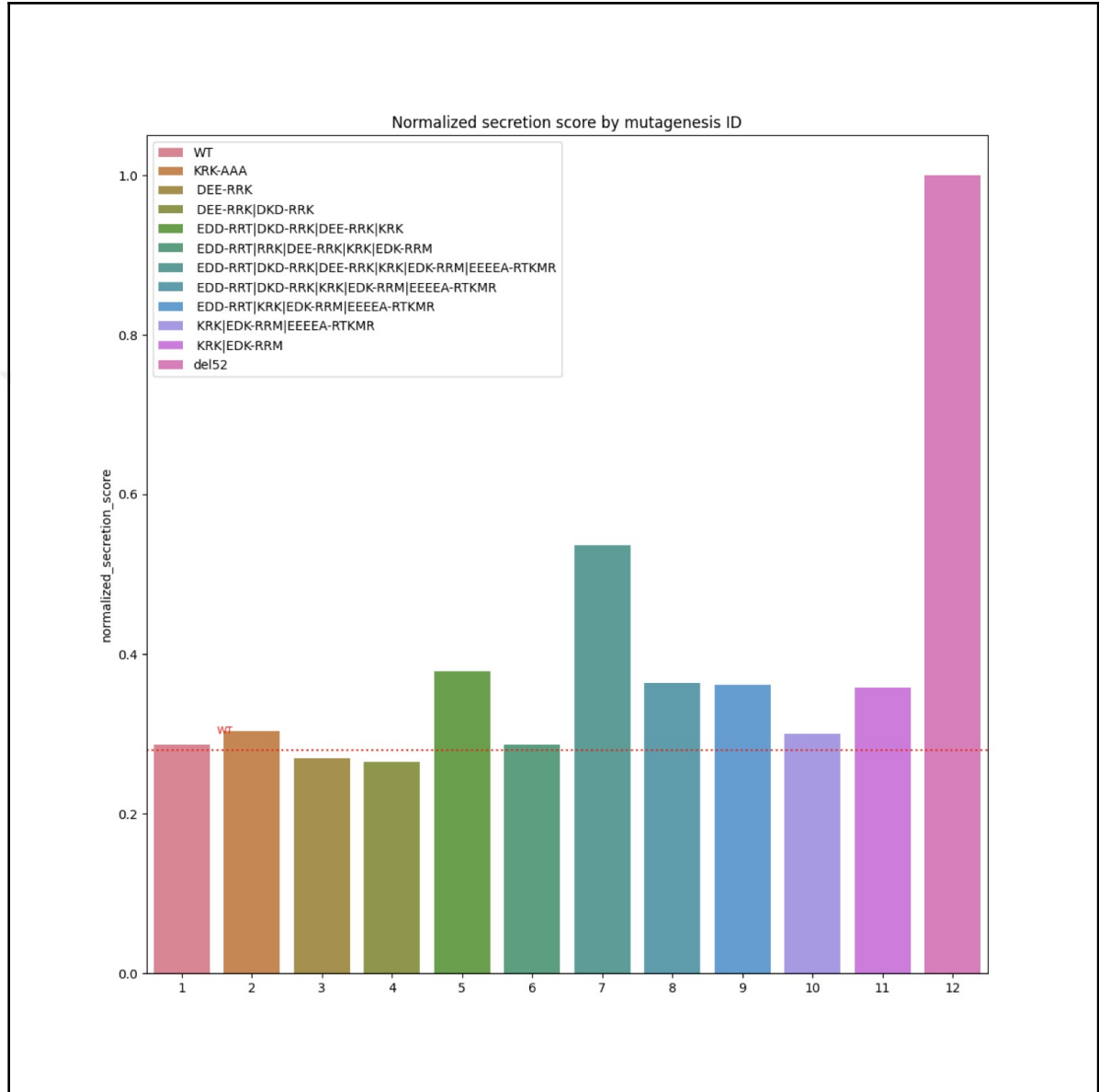


Figure 9d. MULocDeep predicted secretion scores for different CALR WT mutagenesis combinations, with CALR del52 (x=12) used for score normalization. Red dotted line refers to the score for CALR WT.

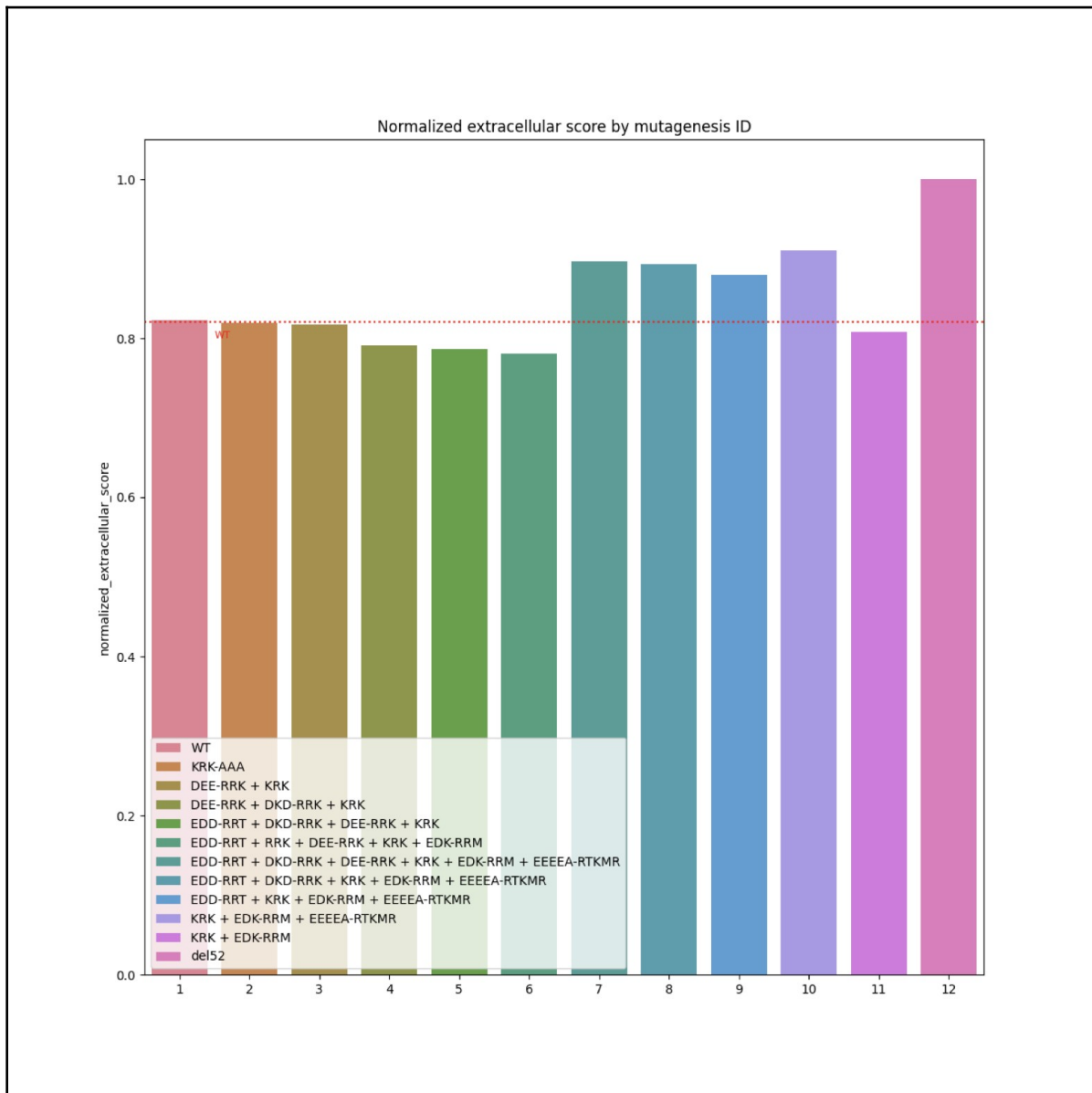


Figure 9e. DeepLoc predicted secretion scores for different CALR WT mutagenesis combinations, with CALR del52 ($\chi=12$) used for score normalization. Red dotted line refers to the score for CALR WT.

3.5) K/R tally gradient approach to predicting levels of CALR del52 secretion

Since the accumulation of positively charged residues through these motifs was consistently contributing to increased secretion, and that mutations to alanines were decreasing predicted secretion, we next assessed whether losing individual lysines or arginines affects secretion. Instead of motifs, a count of lysines and arginines (henceforth referred to as the K/R tally) was utilized for a gradual mutagenesis (See Table 7 as an example) that would permit the use of scatter plots, and thus statistical analyses using the statsmodels.api package in Python. All of these analyses were performed in MULocDeep because of the greater range in predicted values seen in Figures 9a-9e. The values ranged from zero to fifteen (total amount of lysines and arginines in the mutant C-terminus). There were four types of gradients: two types of mutations where lysines and arginines were mutated to either alanines (Figures 10e and 10g) or methionines (Figures 10a and 10c) and two directions in which mutations accumulated – from either left to right or right to left. Methionines were used because in preliminary Western blot data (see Figure 9c), switching arginines to methionines resulted in less CALR del52 secretion. All four gradients showed a consistent trend of less predicted secretion (see Figures 10a, 10c, 10e, and 10g), although the statistical significance (range of R^2 values between 0.85 and 0.92, see Figure 10b, 10d, 10f, and 10h) did slightly differ between them along with the rate of decrease of predicted secretion. Since there was no significant difference in terms of the trend of effects on predicted secretion in terms of direction or type of new amino acids, and also to emulate planned wet-lab mutations, subsequent analyses only used mutations to alanines from the N-terminus to C-terminus.

Table 7. One of the four gradients used to ascertain the effect of losing lysines and arginines on the mutant C-terminus on predicted secretion. In this table, lysines and arginines are gradually mutated to methionines (highlighted in red) from the N-terminus onward. On the top of the table, the CALR del52 C-terminal sequence is on the first row with a “KR_tally” of 15.

KR_tally	sequence
15	TRRMMRTKMRMRMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
14	TMRMMRTKMRMRMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
13	TMMMMRTKMRMRMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
12	TMMMMTKMRMRMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
11	TMMMMMTMRMRMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
10	TMMMMMTMMMRMRMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
9	TMMMMMTMMMRRMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
8	TMMMMMTMMMMRRTRRKMRKMSPARPRTSCREACLQGWTEA
7	TMMMMMTMMMMMRTRRKMRRKMSPARPRTSCREACLQGWTEA
6	TMMMMMTMMMMMMTRRKMRKMSPARPRTSCREACLQGWTEA
5	TMMMMMTMMMMMMTMRKMRRKMSPARPRTSCREACLQGWTEA
4	TMMMMMTMMMMMMTMMKMRKMSPARPRTSCREACLQGWTEA
3	TMMMMMTMMMMMMTMMMRKMSPARPRTSCREACLQGWTEA
2	TMMMMMTMMMMMMTMMMRKMSPARPRTSCREACLQGWTEA
1	TMMMMMTMMMMMMTMMMMKSPARPRTSCREACLQGWTEA
0	TMMMMMTMMMMMMTMMMMMSPARPRTSCREACLQGWTEA

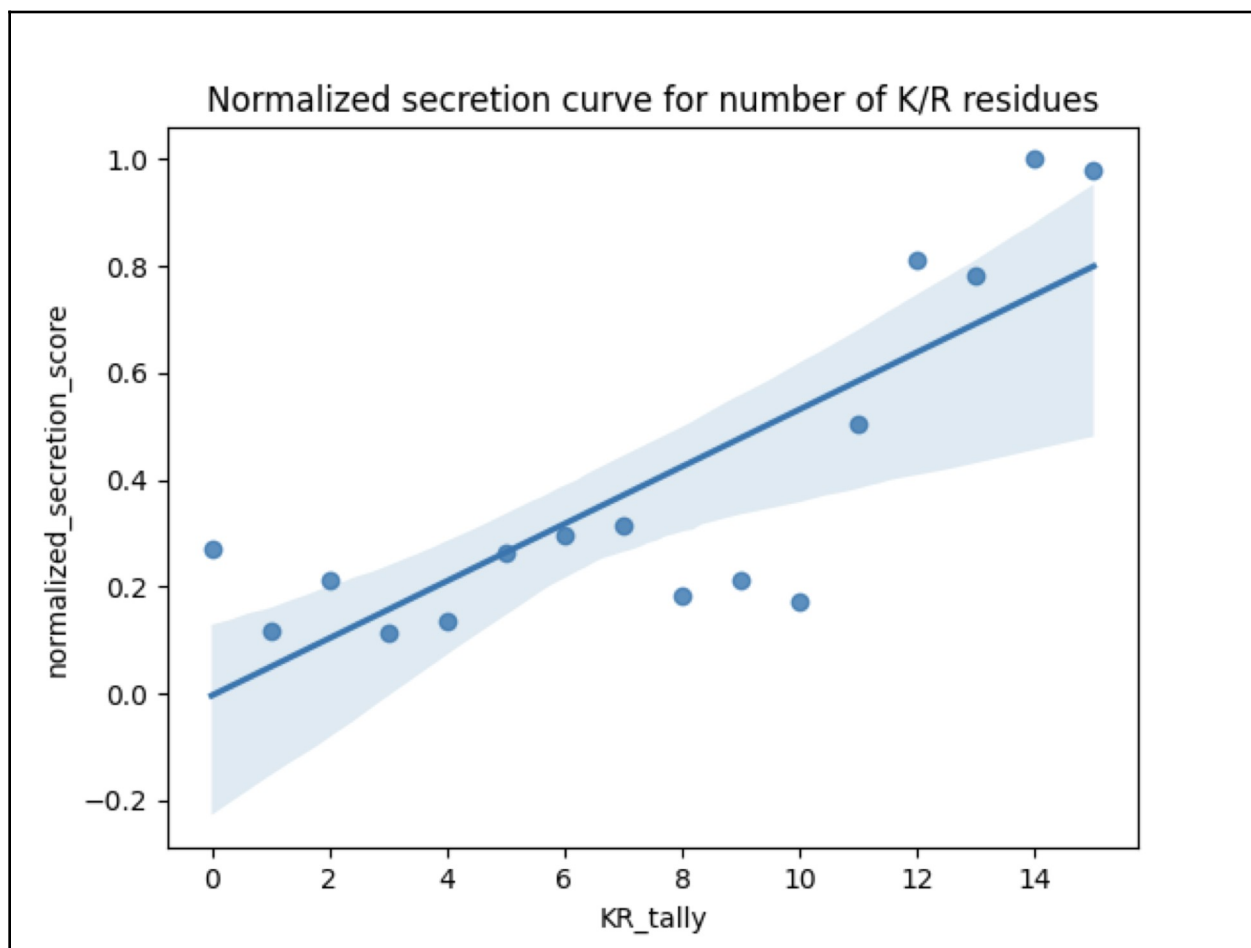


Figure 10a. MULocDeep analysis of the predicted effect on secretion when the K/R tally is mutated to methionines from right to left.

OLS Regression Results						
Dep. Variable:	KR_tally	R-squared (uncentered):	0.874			
Model:	OLS	Adj. R-squared (uncentered):	0.866			
Method:	Least Squares	F-statistic:	104.5			
Date:	Fri	22 Sep 2023	Prob (F-statistic):	3.75e-08		
Time:	15:35:48	Log-Likelihood:	-40.906			
No. Observations:	16	AIC:	83.81			
Df Residuals:	15	BIC:	84.58			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
secreted_loc_score	100.3751	9.821	10.220	0.000	79.442	121.308

Figure 10b. Statistical analyses done using ordinary least-squares regression for data shown in Figure 9a.

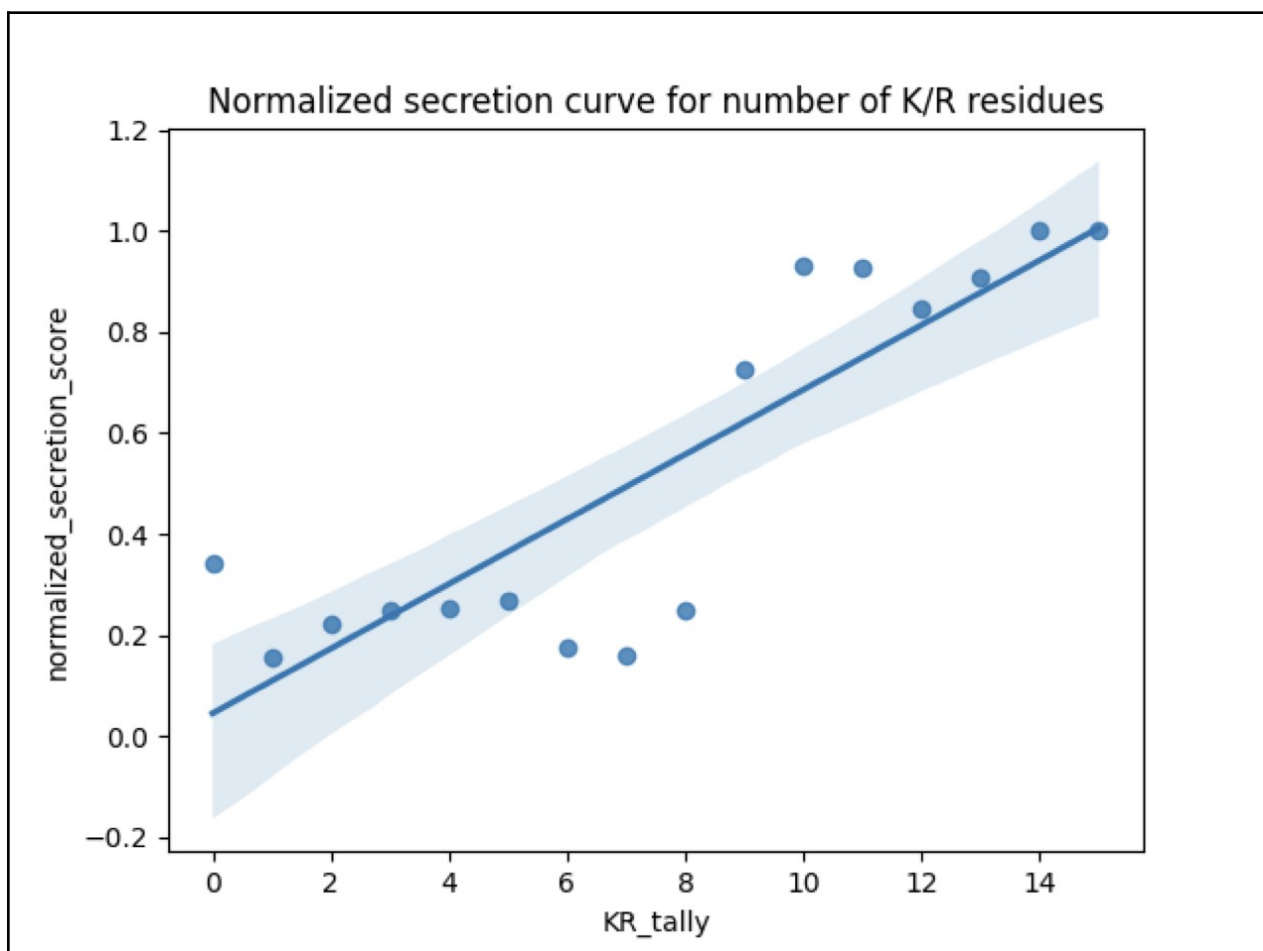


Figure 10c. MULocDeep analysis of the predicted effect on secretion when the K/R tally is mutated to methionines from left to right.

OLS Regression Results						
Dep. Variable:	KR_tally	R-squared (uncentered):	0.922			
Model:	OLS	Adj. R-squared (uncentered):	0.917			
Method:	Least Squares	F-statistic:	177.0			
Date:	Sat	07 Oct 2023	Prob (F-statistic):	1.04e-09		
Time:	12:28:54	Log-Likelihood:	-37.112			
No. Observations:	16	AIC:	76.22			
Df Residuals:	15	BIC:	77.00			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
secreted_loc_score	83.9347	6.310	13.302	0.000	70.486	97.383
Omnibus:	0.246	Durbin-Watson:	0.620			
Prob(Omnibus):	0.884	Jarque-Bera (JB):	0.074			
Skew:	0.129	Prob(JB):	0.964			
Kurtosis:	2.789	Cond. No.	1.00			
Notes:						
[1] R ² is computed without centering (uncentered) since the model does not contain a constant.						
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Figure 10d. Statistical analyses done for data shown in Figure 10c.

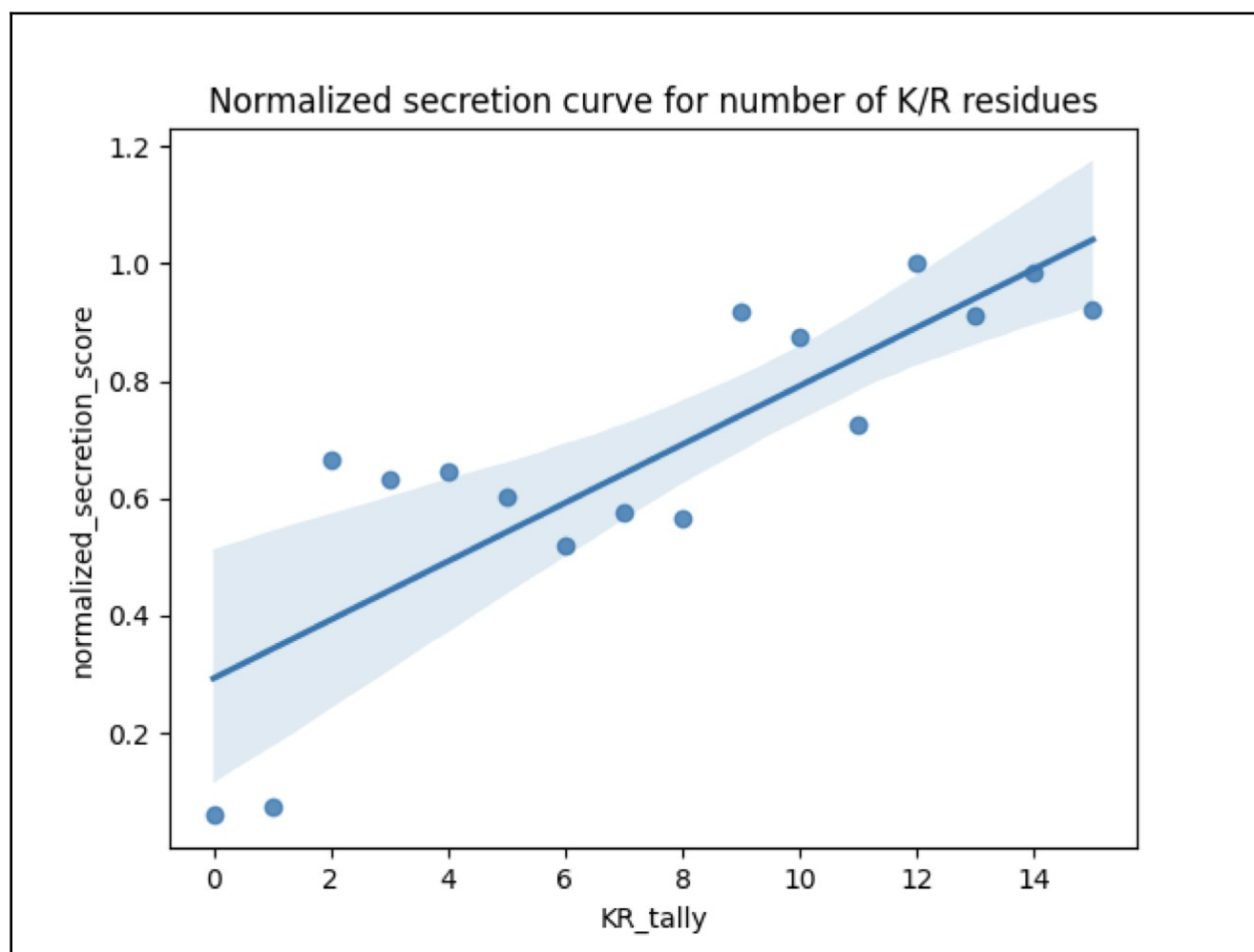


Figure 10e. MULocDeep analysis of the predicted effect on secretion when the K/R tally is mutated to alanines from right to left.

OLS Regression Results						
Dep. Variable:	KR_tally	R-squared (uncentered):	0.911			
Model:	OLS	Adj. R-squared (uncentered):	0.905			
Method:	Least Squares	F-statistic:	152.9			
Date:	Sat	07 Oct 2023	Prob (F-statistic):	2.87e-09		
Time:	12:07:25	Log-Likelihood:	-38.183			
No. Observations:	16	AIC:	78.37			
Df Residuals:	15	BIC:	79.14			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
secreted_loc_score	66.8067	5.403	12.365	0.000	55.291	78.323
Omnibus:	0.650	Durbin-Watson:	0.693			
Prob(Omnibus):	0.723	Jarque-Bera (JB):	0.550			
Skew:	-0.391	Prob(JB):	0.760			
Kurtosis:	2.539	Cond. No.	1.00			
Notes:						
[1] R ² is computed without centering (uncentered) since the model does not contain a constant.						
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Figure 10f. Statistical analyses for data seen in Figure 10e.

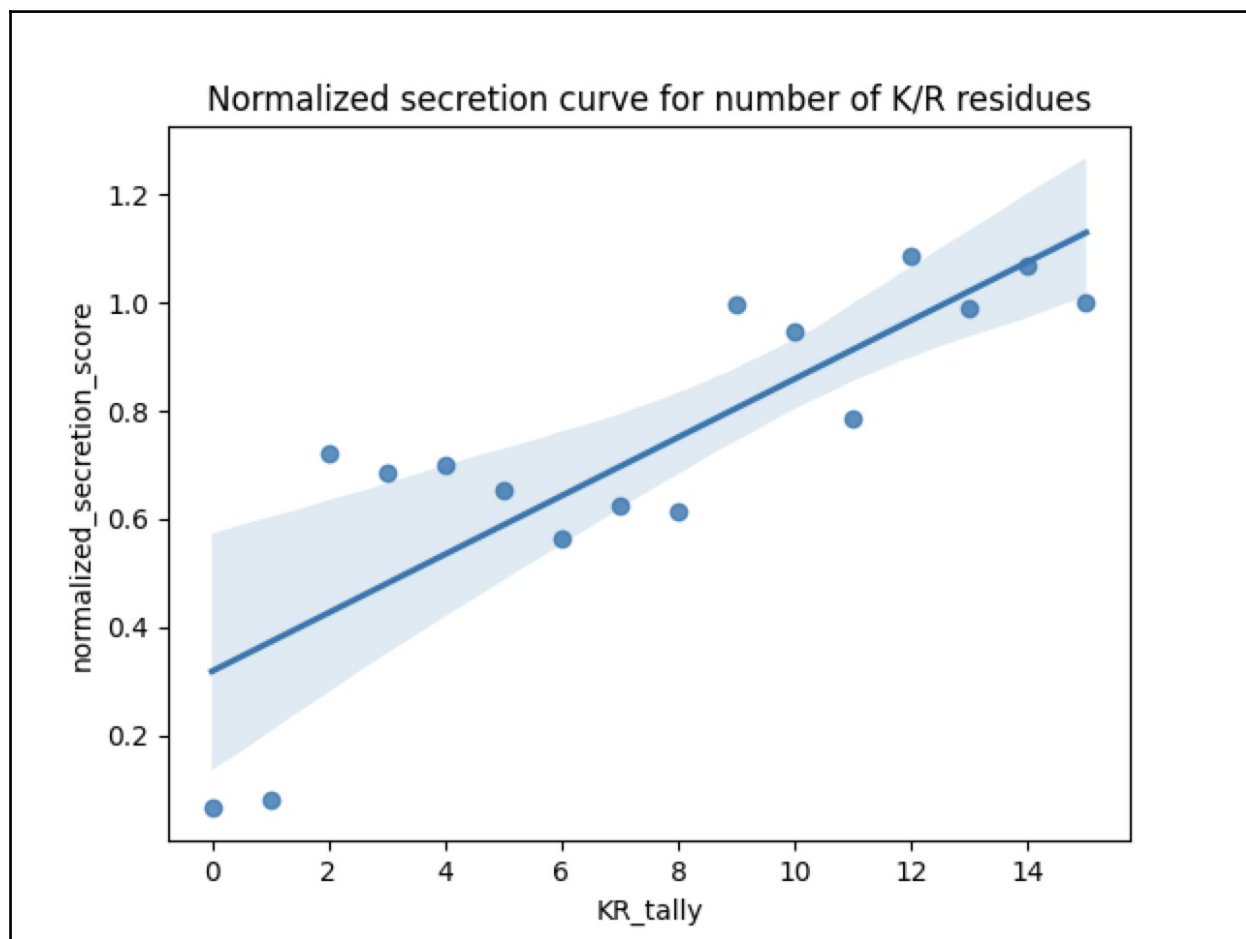


Figure 10g. MULocDeep analysis of the predicted effect on secretion when the K/R tally is mutated to alanines from left to right.

OLS Regression Results						
Dep. Variable:	KR_tally	R-squared (uncentered):	0.911			
Model:	OLS	Adj. R-squared (uncentered):	0.905			
Method:	Least Squares	F-statistic:	152.9			
Date:	Thu	19 Oct 2023	Prob (F-statistic):	2.87e-09		
Time:	07:04:56	Log-Likelihood:	-38.183			
No. Observations:	16	AIC:	78.37			
Df Residuals:	15	BIC:	79.14			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
secreted_loc_score	66.8067	5.403	12.365	0.000	55.291	78.323
Omnibus:	0.650	Durbin-Watson:	0.693			
Prob(Omnibus):	0.723	Jarque-Bera (JB):	0.550			
Skew:	-0.391	Prob(JB):	0.760			
Kurtosis:	2.539	Cond. No.	1.00			
Notes:						
[1] R ² is computed without centering (uncentered) since the model does not contain a constant.						
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Figure 10h. Statistical analyses for data seen in Figure 9g.

3.6) Addition of K/R residues to CALR del52 towards either terminus

Since the statsmodels.api package recommends at least 20 data points, and also to assess whether increasing the number of lysines and arginines would proportionally augment predicted secretion levels, residues on either side of the proposed motifs, towards either the C-terminus or the N-terminus, were mutated to either lysines or arginines. Up to seven residues (see Tables 8a-8b) were substituted this way and these new sequences were then processed with either MULocDeep or DeepLoc. The results show no proportional increase, but rather either stagnation or a slight decline (see Figures 11a-11d) in terms of predicted secretion, suggesting that it is possible to oversaturate the terminus with these motifs to the point where it could hinder the ability of mutant CALR to move outside the ER.

Table 8a. One of the lists of sequences used for “extra K/R” analysis – additions (in blue) on the N-terminal side. Alanines that substituted lysines and arginines highlighted in red.

KR_tally	sequence
22	DKRRRRRRRRMMRTKMRMRMRRTRRKMRRKMSPARPTSCREACLQGWTEA
21	DKQRRRRRRRRMMRTKMRMRMRRTRRKMRRKMSPARPTSCREACLQGWTEA
20	DKQDRRRRRRRRMMRTKMRMRMRRTRRKMRRKMSPARPTSCREACLQGWTEA
19	DKQDERRRRRRMMRTKMRMRMRRTRRKMRRKMSPARPTSCREACLQGWTEA
18	DKQDEERRRRRMMRTKMRMRMRRTRRKMRRKMSPARPTSCREACLQGWTEA
17	DKQDEEQRRRRMMRTKMRMRMRRTRRKMRRKMSPARPTSCREACLQGWTEA
16	DKQDEEQRRRRMMRTKMRMRMRRTRRKMRRKMSPARPTSCREACLQGWTEA
15	DKQDEEQRTRRMMRTKMRMRMRRTRRKMRRKMSPARPTSCREACLQGWTEA
14	DKQDEEQRTARRMMRTKMRMRMRRTRRKMRRKMSPARPTSCREACLQGWTEA
13	DKQDEEQRTAAMMRTKMRMRMRRTRRKMRRKMSPARPTSCREACLQGWTEA
12	DKQDEEQRTAAMMATKMRMRMRRTRRKMRRKMSPARPTSCREACLQGWTEA
11	DKQDEEQRTAAMMATAMRMRMRRTRRKMRRKMSPARPTSCREACLQGWTEA
10	DKQDEEQRTAAMMATAMAMRMRMRRTRRKMRRKMSPARPTSCREACLQGWTEA
9	DKQDEEQRTAAMMATAMAMARMRRTRRKMRRKMSPARPTSCREACLQGWTEA
8	DKQDEEQRTAAMMATAMAMAMRRTRRKMRRKMSPARPTSCREACLQGWTEA
7	DKQDEEQRTAAMMATAMAMAMARTRRKMRRKMSPARPTSCREACLQGWTEA
6	DKQDEEQRTAAMMATAMAMAMATRRKMRRKMSPARPTSCREACLQGWTEA
5	DKQDEEQRTAAMMATAMAMAMATAARKMRRRKMSPARPTSCREACLQGWTEA
4	DKQDEEQRTAAMMATAMAMAMATAAKMRRRKMSPARPTSCREACLQGWTEA
3	DKQDEEQRTAAMMATAMAMAMATAAAMRRKMSPARPTSCREACLQGWTEA
2	DKQDEEQRTAAMMATAMAMAMATAAAMARKMSPARPTSCREACLQGWTEA
1	DKQDEEQRTAAMMATAMAMAMATAAAMAAKMSPARPTSCREACLQGWTEA
0	DKQDEEQRTAAMMATAMAMAMATAAAMAAAMSPARPTSCREACLQGWTEA

Table 8b. This list of sequences uses the same “K/R tally” as before (alanines substituting lysines and arginines shown in red), but also has lysines and arginines substituting residues towards the C-terminus, which are highlighted in blue.

KR_tally	sequence
22	TRMMRTKMRMRMRRTRRKMRRKRRRRRKRRKCREACLQGWTEA
21	TRMMRTKMRMRMRRTRRKMRRKRRRRRKRRSCREACLQGWTEA
20	TRMMRTKMRMRMRRTRRKMRRKRRRRRKRTSCREACLQGWTEA
19	TRMMRTKMRMRMRRTRRKMRRKRRRRRPRTSCREACLQGWTEA
18	TRMMRTKMRMRMRRTRRKMRRKRRRARPRTSCREACLQGWTEA
17	TRMMRTKMRMRMRRTRRKMRRKRRPARPRTSCREACLQGWTEA
16	TRMMRTKMRMRMRRTRRKMRRKRSARPRTSCREACLQGWTEA
15	TRMMRTKMRMRMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
14	TARMRTKMRMRMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
13	TAAMMRTKMRMRMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
12	TAAMMATKMRMRMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
11	TAAMMATAMRMRMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
10	TAAMMATAMAMRMRMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
9	TAAMMATAMAMARMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
8	TAAMMATAMAMAAMRRTRRKMRRKMSPARPRTSCREACLQGWTEA
7	TAAMMATAMAMAAMARTRRKMRRKMSPARPRTSCREACLQGWTEA
6	TAAMMATAMAMAAMAATRRKMRRKMSPARPRTSCREACLQGWTEA
5	TAAMMATAMAMAAMAATARKMRKMSPARPRTSCREACLQGWTEA
4	TAAMMATAMAMAAMAATAAKMRKMSPARPRTSCREACLQGWTEA
3	TAAMMATAMAMAAMAATAAAMRRKMSPARPRTSCREACLQGWTEA
2	TAAMMATAMAMAAMAATAAAMARKMSPARPRTSCREACLQGWTEA
1	TAAMMATAMAMAAMAATAAAMAAMKSPARPRTSCREACLQGWTEA
0	TAAMMATAMAMAAMAATAAAMAAMSPARPRTSCREACLQGWTEA

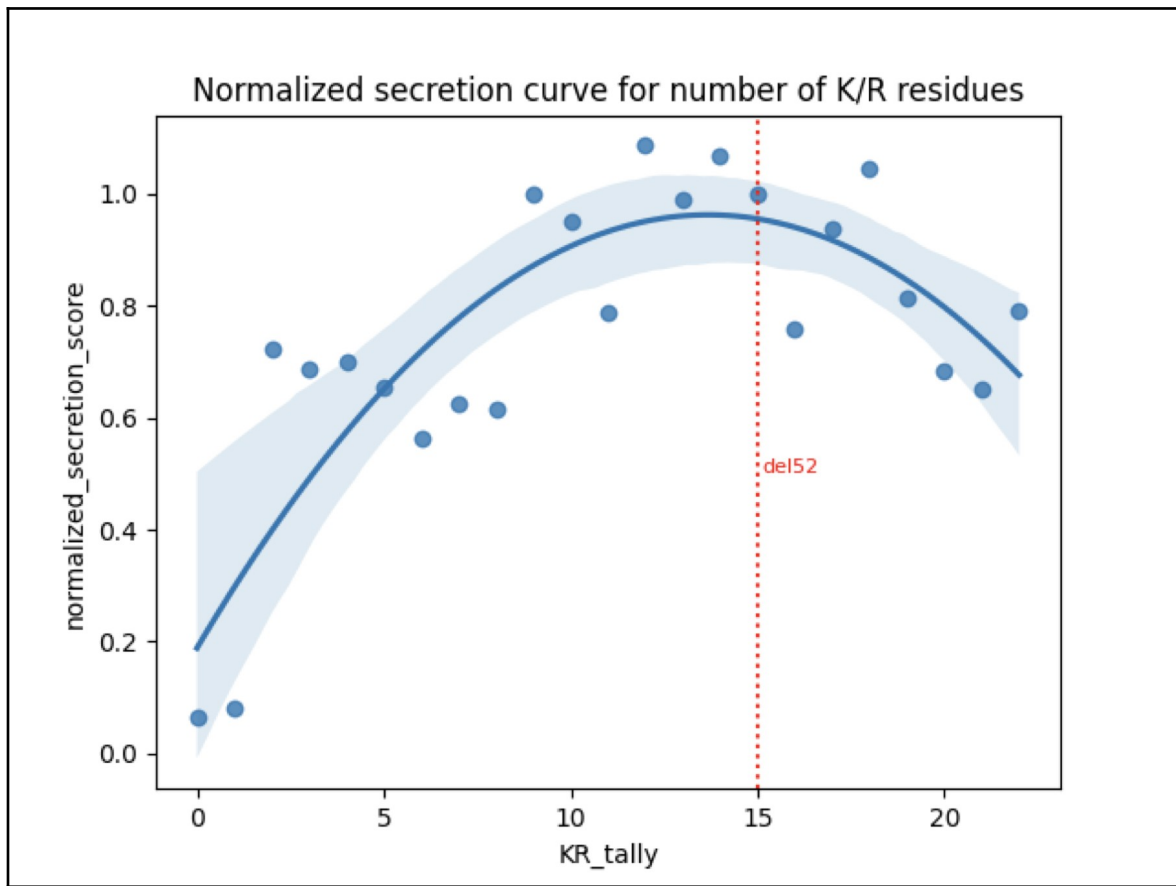


Figure 11a. MULocDeep predicted secretion scores normalized to CALR del52 (shown with dotted red line with KR_tally value of 15) with the additional lysines or arginines towards the N-terminus.

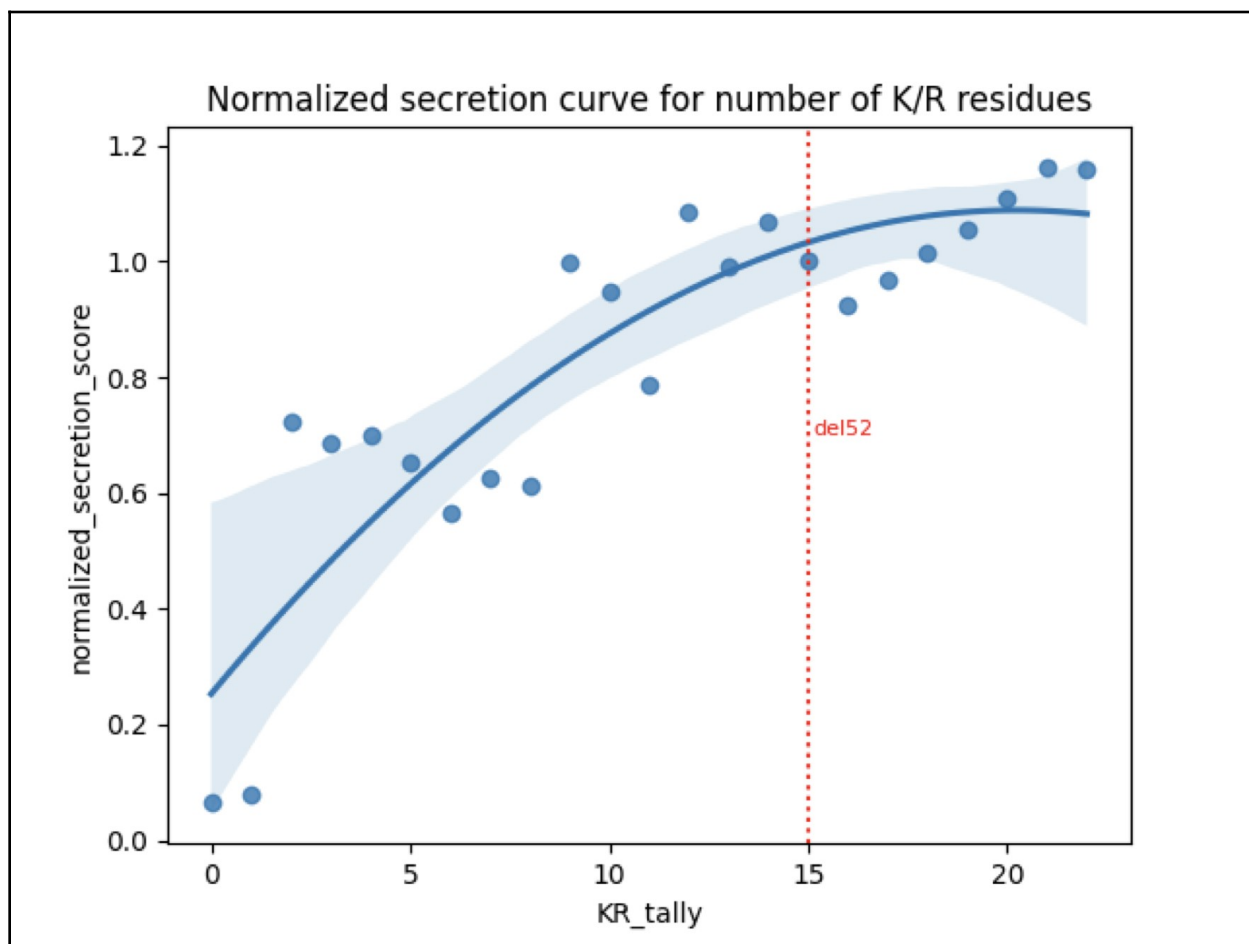


Figure 11b. MULocDeep predicted secretion scores normalized to CALR del52 (shown with dotted red line with KR_tally value of 15) with the additional lysines or arginines in the C-terminal direction.

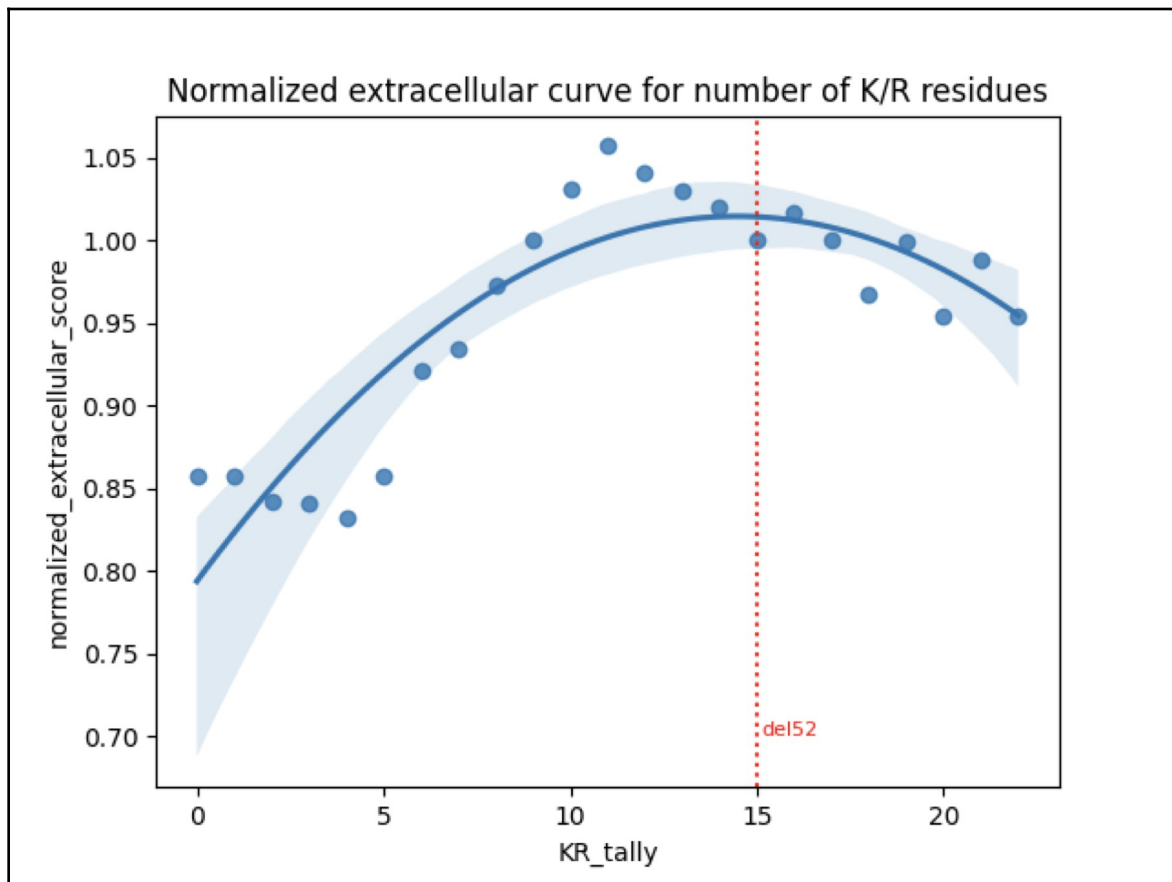


Figure 11c. DeepLoc predicted secretion scores normalized to CALR del52 (shown with dotted red line with KR_tally value of 15) with the additional lysines or arginines towards the N-terminus.

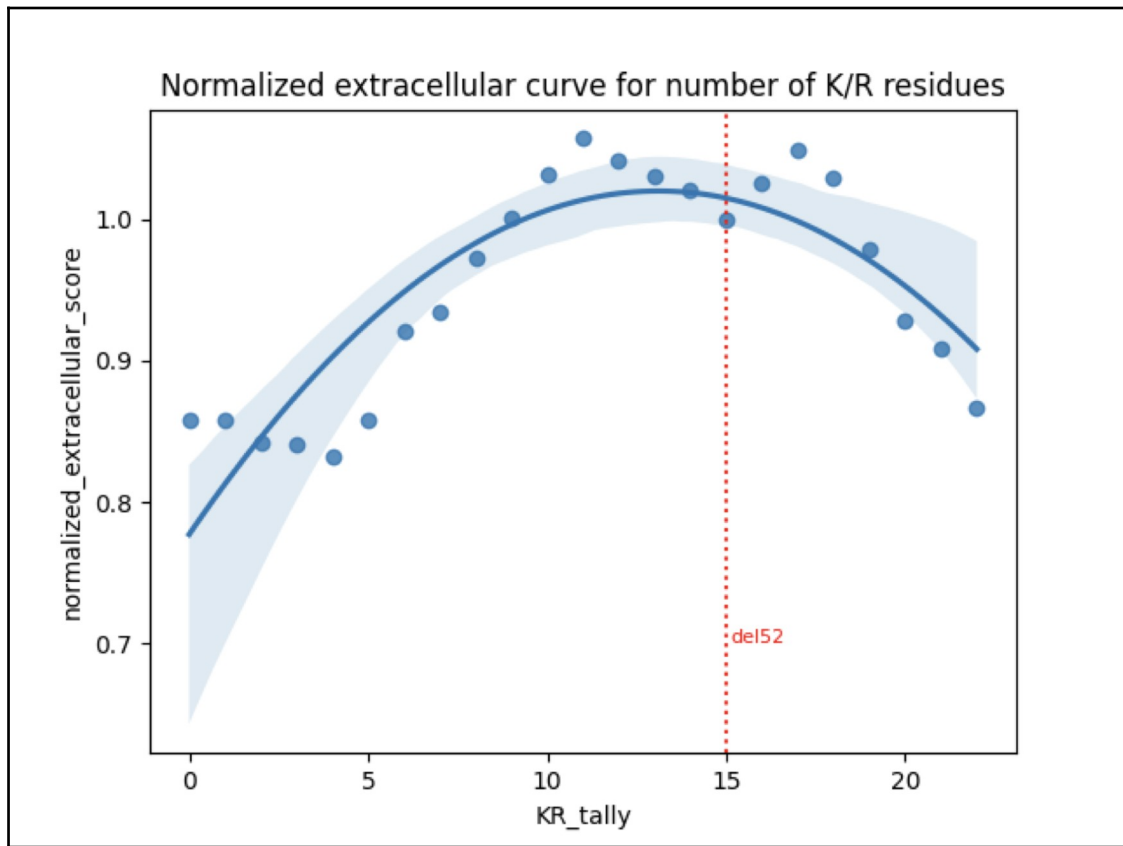


Figure 11d. DeepLoc predicted secretion scores normalized to CALR del52 (shown with dotted red line with KR_tally value of 15) with the additional lysines or arginines in the C-terminal direction.

OLS Regression Results						
Dep. Variable:	KR_tally	R-squared (uncentered):	0.909			
Model:	OLS	Adj. R-squared (uncentered):	0.905			
Method:	Least Squares	F-statistic:	219.4			
Date:	Mon	16 Oct 2023	Prob (F-statistic):	6.33e-13		
Time:	06:53:59	Log-Likelihood:	-63.807			
No. Observations:	23	AIC:	129.6			
Df Residuals:	22	BIC:	130.7			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
secreted_loc_score	87.0394	5.876	14.812	0.000	74.853	99.226
Omnibus:	1.193	Durbin-Watson:	0.387			
Prob(Omnibus):	0.551	Jarque-Bera (JB):	0.862			
Skew:	-0.102	Prob(JB):	0.650			
Kurtosis:	2.074	Cond. No.	1.00			
Notes:						
[1] R ² is computed without centering (uncentered) since the model does not contain a constant.						
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Figure 11e. Statistical analyses for data in Figure 11a.

OLS Regression Results						
Dep. Variable:	KR_tally	R-squared (uncentered):	0.817			
Model:	OLS	Adj. R-squared (uncentered):	0.809			
Method:	Least Squares	F-statistic:	98.25			
Date:	Sun	15 Oct 2023	Prob (F-statistic):	1.42e-09		
Time:	13:29:51	Log-Likelihood:	-71.821			
No. Observations:	23	AIC:	145.6			
Df Residuals:	22	BIC:	146.8			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
secreted_loc_score	90.9769	9.178	9.912	0.000	71.942	110.012
Omnibus:	2.227	Durbin-Watson:	0.279			
Prob(Omnibus):	0.328	Jarque-Bera (JB):	1.848			
Skew:	0.659	Prob(JB):	0.397			
Kurtosis:	2.560	Cond. No.	1.00			
Notes:						
[1] R ² is computed without centering (uncentered) since the model does not contain a constant.						
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Figure 11f. Statistical analyses for data seen in Figure 11b.

OLS Regression Results						
Dep. Variable:	KR_tally	R-squared (uncentered):	0.775			
Model:	OLS	Adj. R-squared (uncentered):	0.764			
Method:	Least Squares	F-statistic:	75.59			
Date:	Mon	16 Oct 2023	Prob (F-statistic):	1.44e-08		
Time:	06:34:19	Log-Likelihood:	-74.222			
No. Observations:	23	AIC:	150.4			
Df Residuals:	22	BIC:	151.6			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
extracellular_loc_score	25.3993	2.921	8.694	0.000	19.341	31.458
Omnibus:	2.779	Durbin-Watson:	0.026			
Prob(Omnibus):	0.249	Jarque-Bera (JB):	1.360			
Skew:	0.215	Prob(JB):	0.507			
Kurtosis:	1.889	Cond. No.	1.00			
Notes:						
[1] R ² is computed without centering (uncentered) since the model does not contain a constant.						
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Figure 11g. Statistical analyses for data seen in Figure 11c.

OLS Regression Results						
Dep. Variable:	KR_tally	R-squared (uncentered):	0.765			
Model:	OLS	Adj. R-squared (uncentered):	0.755			
Method:	Least Squares	F-statistic:	71.78			
Date:	Sun	15 Oct 2023	Prob (F-statistic):	2.25e-08		
Time:	13:39:21	Log-Likelihood:	-74.681			
No. Observations:	23	AIC:	151.4			
Df Residuals:	22	BIC:	152.5			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
extracellular_loc_score	25.3424	2.991	8.472	0.000	19.139	31.546
Omnibus:	1.869	Durbin-Watson:	0.027			
Prob(Omnibus):	0.393	Jarque-Bera (JB):	1.240			
Skew:	0.299	Prob(JB):	0.538			
Kurtosis:	2.033	Cond. No.	1.00			
Notes:						
[1] R ² is computed without centering (uncentered) since the model does not contain a constant.						
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Figure 11h. Statistical analyses done for data seen in 11d.

3.7) Examination of KKRK motif in the CALR WT sequence

Although the statistical correlations between the K/R tally and the predicted secretion according to either MULocDeep or DeepLoc were still significant (>0.8 , see Figures 11e-11h), these values were still lower than that of the previous data due to predicted secretion either plateauing for N-terminal additions or declining in the case of C-terminal additions. These data suggest that these motifs may enable mutant CALR secretion up to a certain point, and also their specific location on the sequence is optimal for enhancing secretion. If these motifs were either further upstream or downstream of their current location, there would possibly not be the same effect of enhancing secretion. The main consensus of the accumulated predictions so far suggests that these positively charged motifs play a role in enhancing secretion, albeit as long as there are not too many of them. Conversely, to see if removing negatively charged residues also promoted secretion, almost a dozen aspartic and glutamic acids flanking the KKRK residues at positions 374-377 in WT CALR (see Table 9a) were mutated to either alanines or lysines / arginines. Additionally, the KKRK motif was screened in the ELM search engine both in the canonical sequence and when the adjacent residues were mutated to alanines (see Table 9b), based on the higher increase in predicted secretion when the motifs flanking the KKRK residues (see Table 6 and Figures 9a-9b). This possible motif did not surpass the threshold set by ELM with the canonical WT sequence but it did once the nearby residues were substituted with alanines. Using the DeepLoc sorting signal features (see Figure 12a), mutating nearby residues to alanines did not change the predicted significance of the KKRK motif – in fact, it still remained at zero, but adding extra arginines and lysines where the aspartic and glutamic acids were located did make a difference (Figure 12b).

Table 9a. Screenshot of search results of CALR WT in the ELM search engine. KKRK is implicated to be a possible motif, but due to the nature of nearby residues, is unable to function as such due to lack of exposure.

■ **List of excluded ELMs falling inside SMART/PFAM domains and/or scoring poorly with the structural filter (if applicable).**

Matches in this list are only likely to be of interest if they are in accessible surface-exposed loops. Motif matches buried in stably folded cores of globular domains are not plausible candidates.
If the [structural filter](#) (BETA version) is applicable it is possible to view these structures with [Jmol](#). For more info consult the [PDB](#) structure entry used for structure filtering or the [SMART](#) or [PFAM](#) entries for useful links to solved 3D structures.

Elm Name	Positions	View in Jmol	Elm Description	Cell Compartment	Pattern	PHI-Blast Instance Mapping	Structural Filter Info	Probability
CLV_NRD_NRD_1	375-377 [A]	–	N-Arg dibasic convertase (NRD/Nardilysin) cleavage site (X- -R-K or R- -R-X).	extracellular, Golgi apparatus, cell surface	(.RK) (RR[^(KR)]	–	–	7.465e-03
CLV_PCSK_KEX2_1	375-377 [A]	–	Yeast kexin 2 cleavage site (K-R- -X or R-R- -X).	extracellular, Golgi apparatus	[KR]R.	–	–	7.973e-03
CLV_PCSK_PC1ET2_1	375-377 [A]	–	NEC1/NEC2 cleavage site (K-R- -X).	extracellular, Golgi apparatus, Golgi membrane	KR.	–	–	3.903e-03

Table 9b. ELM search results for KRK motif when nearby residues are mutated to alanines. Color of the table changes because the threshold has been surpassed, meaning that this motif is likely to be active due to newly gained exposure.

CLV_NRD_NRD_1	KRK	375-377 [A]	–	N-Arg dibasic convertase (NRD/Nardilysin) cleavage site (X- -R-K or R- -R-X).	extracellular, Golgi apparatus, cell surface
CLV_PCSK_KEX2_1	KRK	375-377 [A]	–	Yeast kexin 2 cleavage site (K-R- -X or R-R- -X).	extracellular, Golgi apparatus
CLV_PCSK_PC1ET2_1	KRK	375-377 [A]	–	NEC1/NEC2 cleavage site (K-R- -X).	extracellular, Golgi apparatus, Golgi membrane

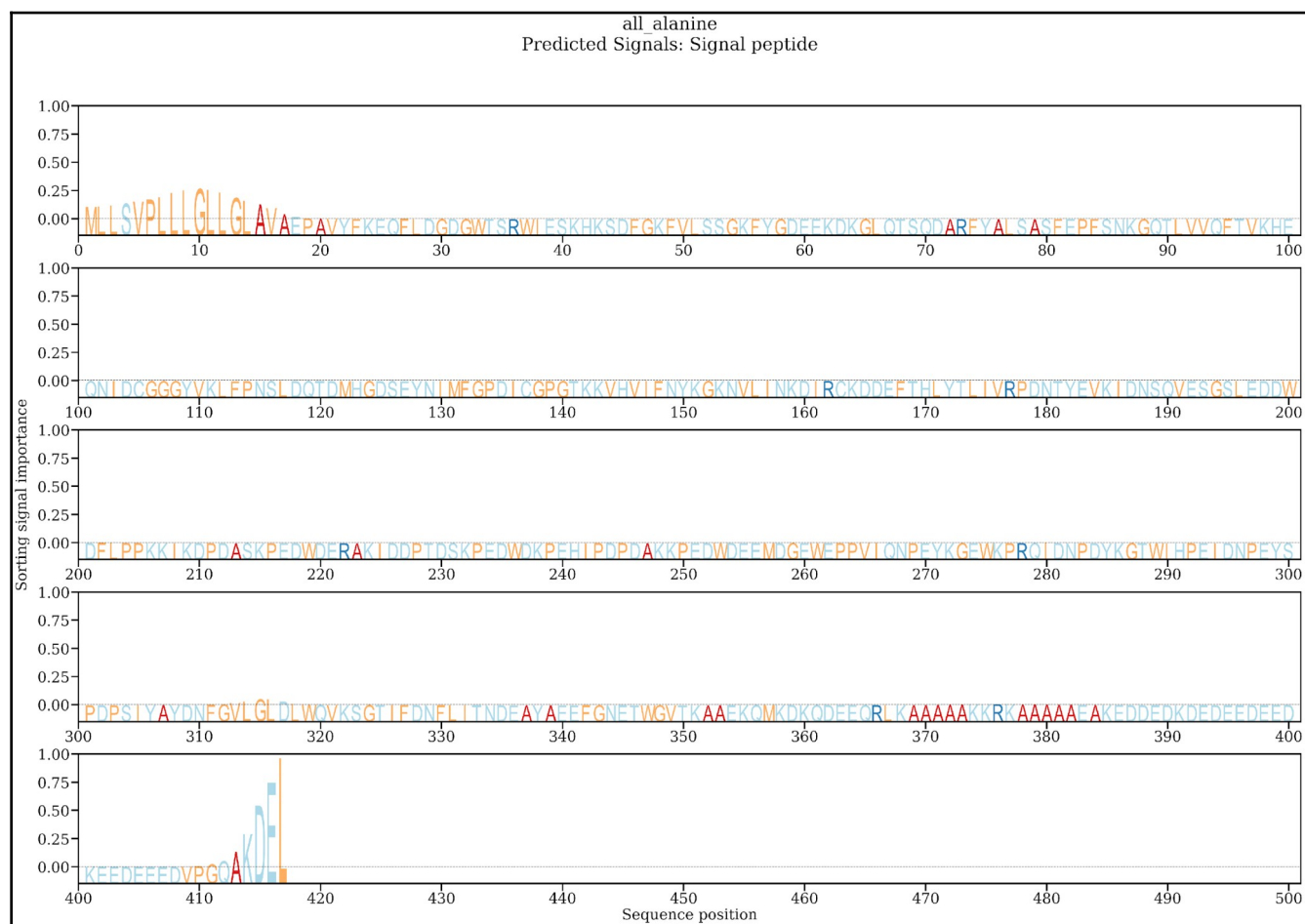


Figure 12a. DeepLoc sorting signal graph for CALR WT with select residues around KKRK mutated to alanines – y axis corresponds to predicted importance for sorting signals (see KDEL anchor at the end).

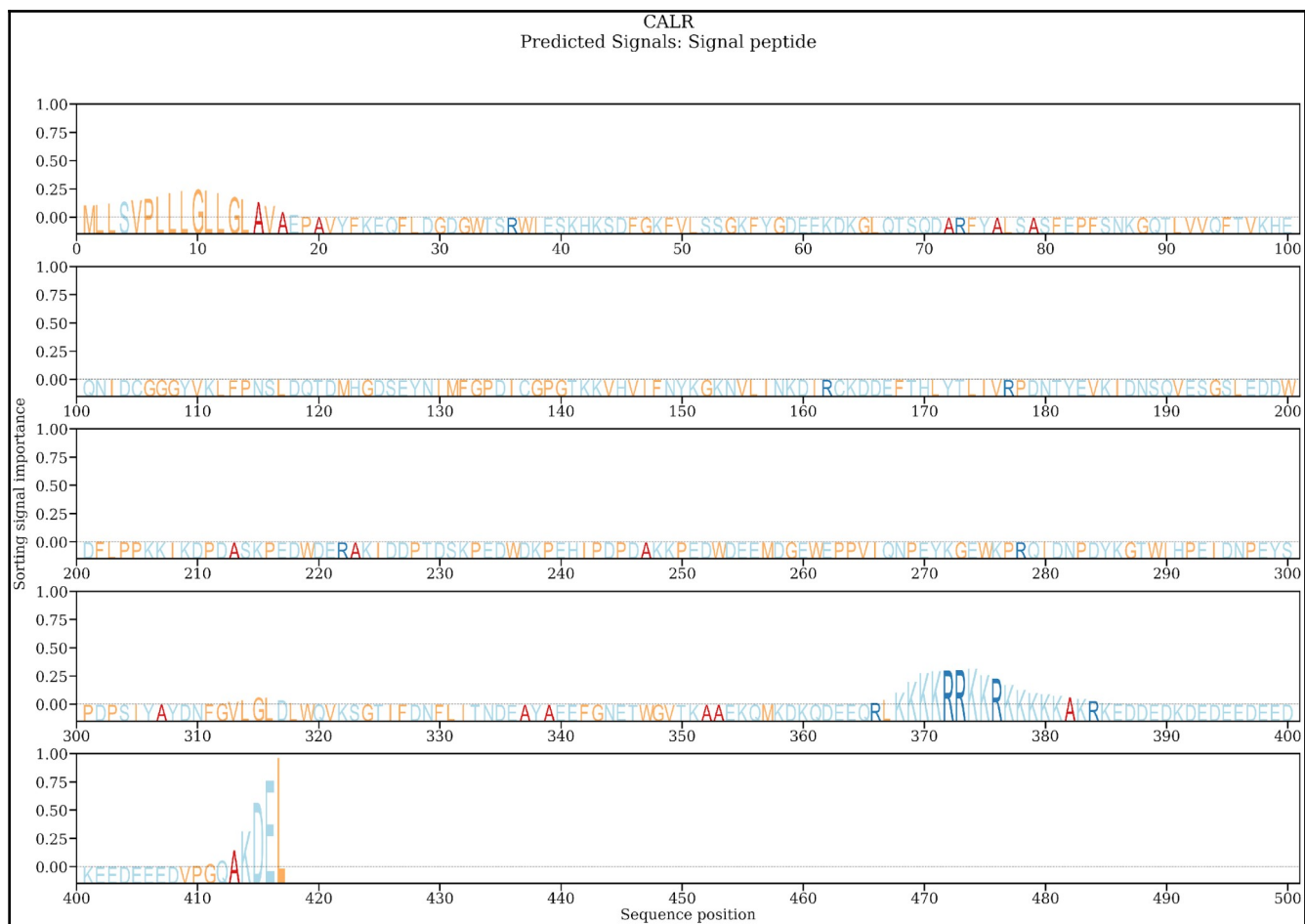


Figure 12b. DeepLoc sorting signal graph using modified sequence where in the vicinity of KKRK, aspartic acids were mutated to lysines and aspartic acids were mutated to arginines. Heightened sorting signal values are observed between positions 370 to 380.

3.8) Visualization of CALR PDB files with PyMOL

To further understand how the loss of aspartic and glutamic acids affects sub-cellular localization, a PDB file for CALR WT was obtained using LocalColabFold and then these residues were mutated to alanines in PyMOL (see Figure 13c-13d) similar to what was done for the proposed secretion motifs in CALR del52 (see Table 3c for reference). Distances in angstroms between residues were recorded (see yellow dotted lines on Figure 13a) using the measurement tool provided by PyMOL. As seen in Figure 13a, the side chains of the KKRK motif are likely to interact with nearby oppositely charged acidic residues as the distances are relatively short ($< 8 \text{ \AA}$), which likely explains why this motif did not appear in the WT or ins5 DeepLoc sorting signal results (see Figures 7a and 7c). Generally, the KKRK motif was far more open and exposed as the alanine side chains were much further away (see Figure 13d), thus accounting for the differences in the ELM search motif results (Tables 9a-9b). This result is in agreement with the ELM search result explanation for the change in likelihood for KKRK acting as a motif with the modified sequence (Table 9b).

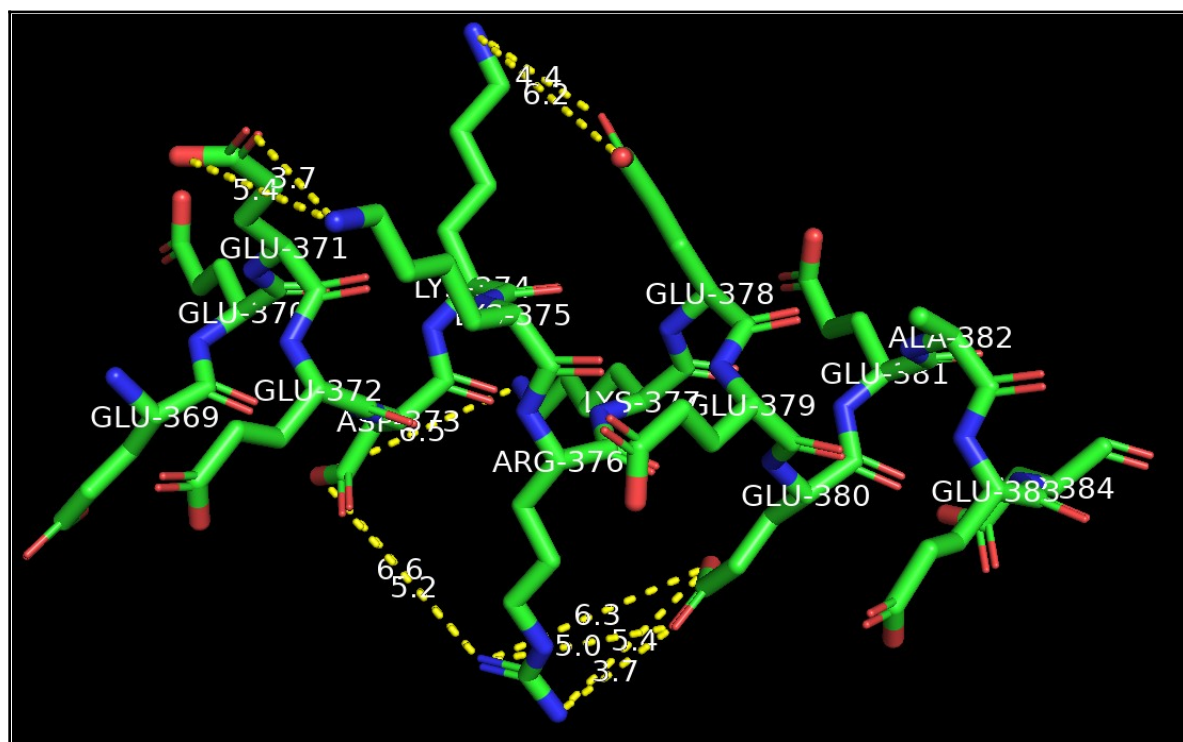


Figure 13a. PyMOL-generated snapshot of residues surrounding KKRK in CALR WT. Yellow dotted lines represent distances between atoms of opposite charges (red for oxygen on aspartic acids or glutamic acids and blue for nitrogens on lysines or arginines). Numbers are in angstroms. Color scheme: green for carbons, red for oxygens, and blue for nitrogens.

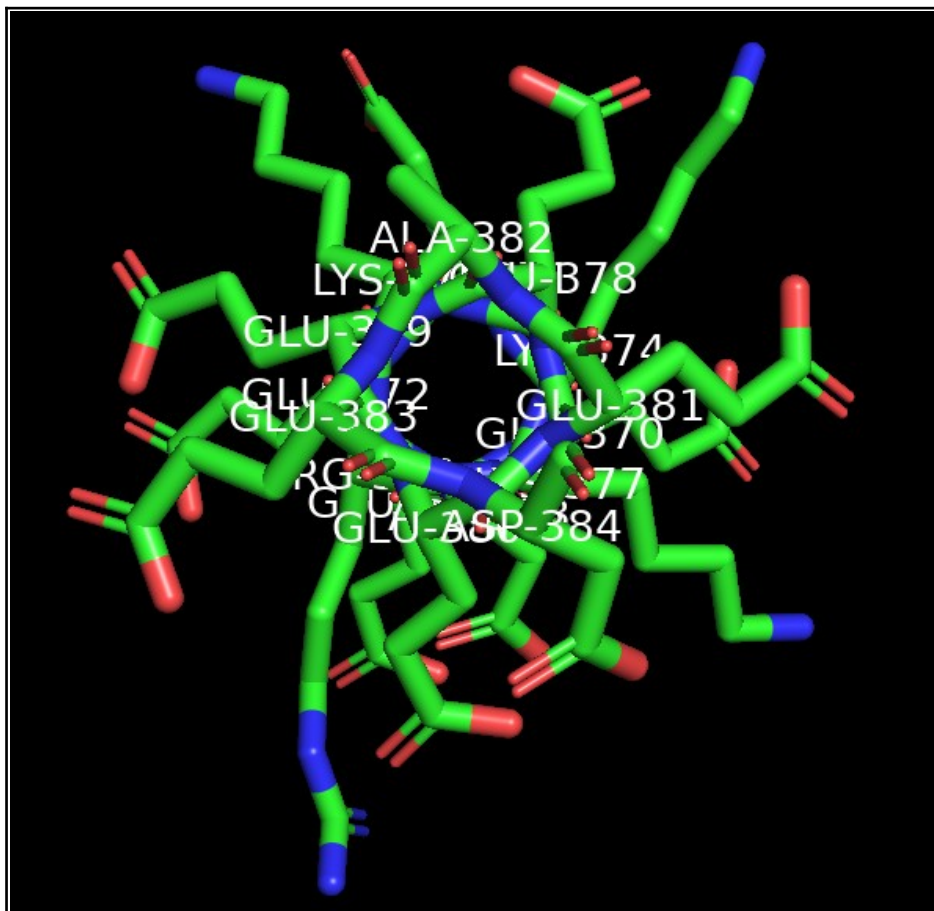


Figure 13b. Side view of residues shown in 13a to highlight potential interactions between amino acids with oppositely charged side chains.

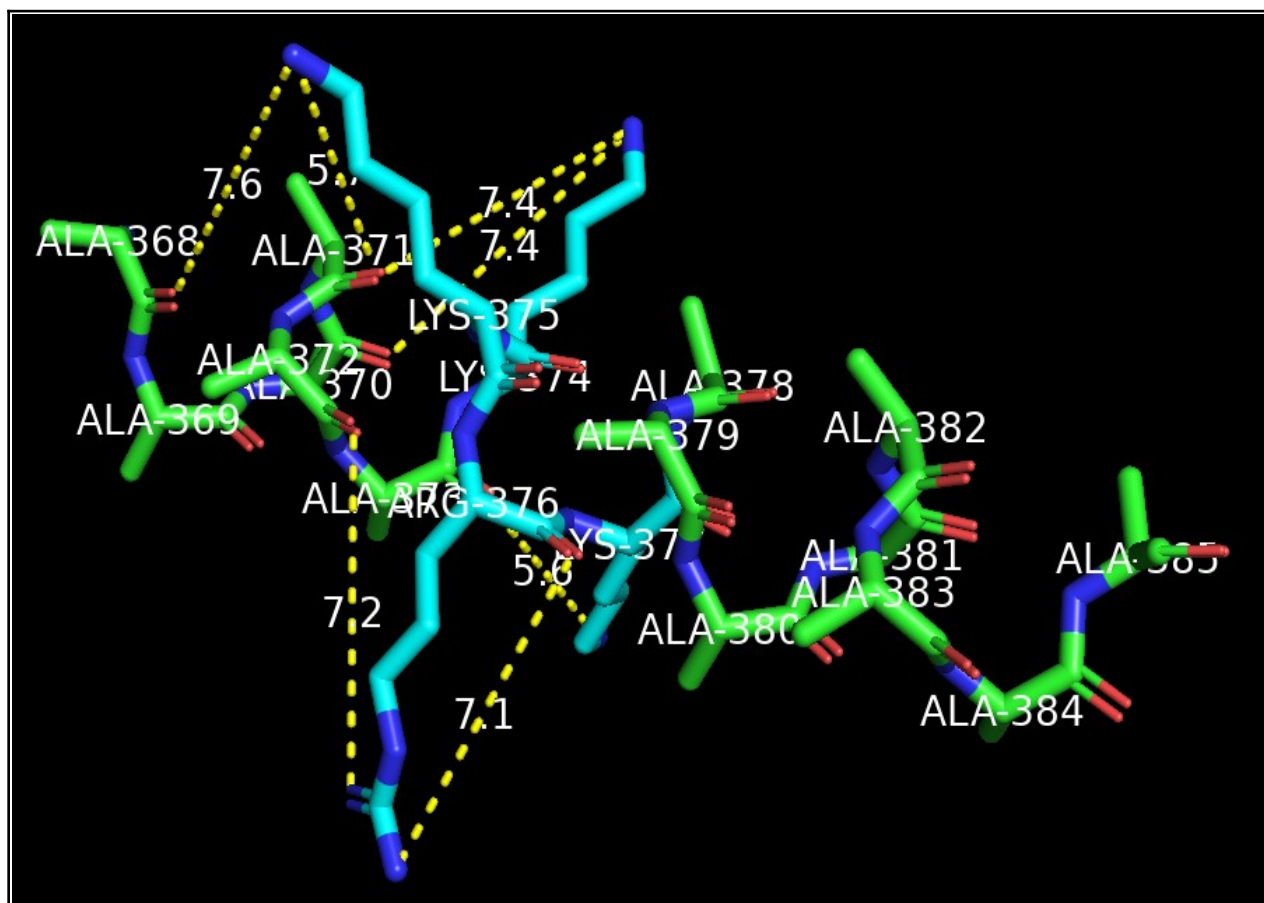


Figure 13c. PyMOL-generated snapshot of potential interactions when KKRK (highlighted in teal for carbons rather than green, nitrogens are still blue and oxygens are still red) adjacent residues are all mutated to alanines. Dotted lines represent distances between select atoms in angstroms.

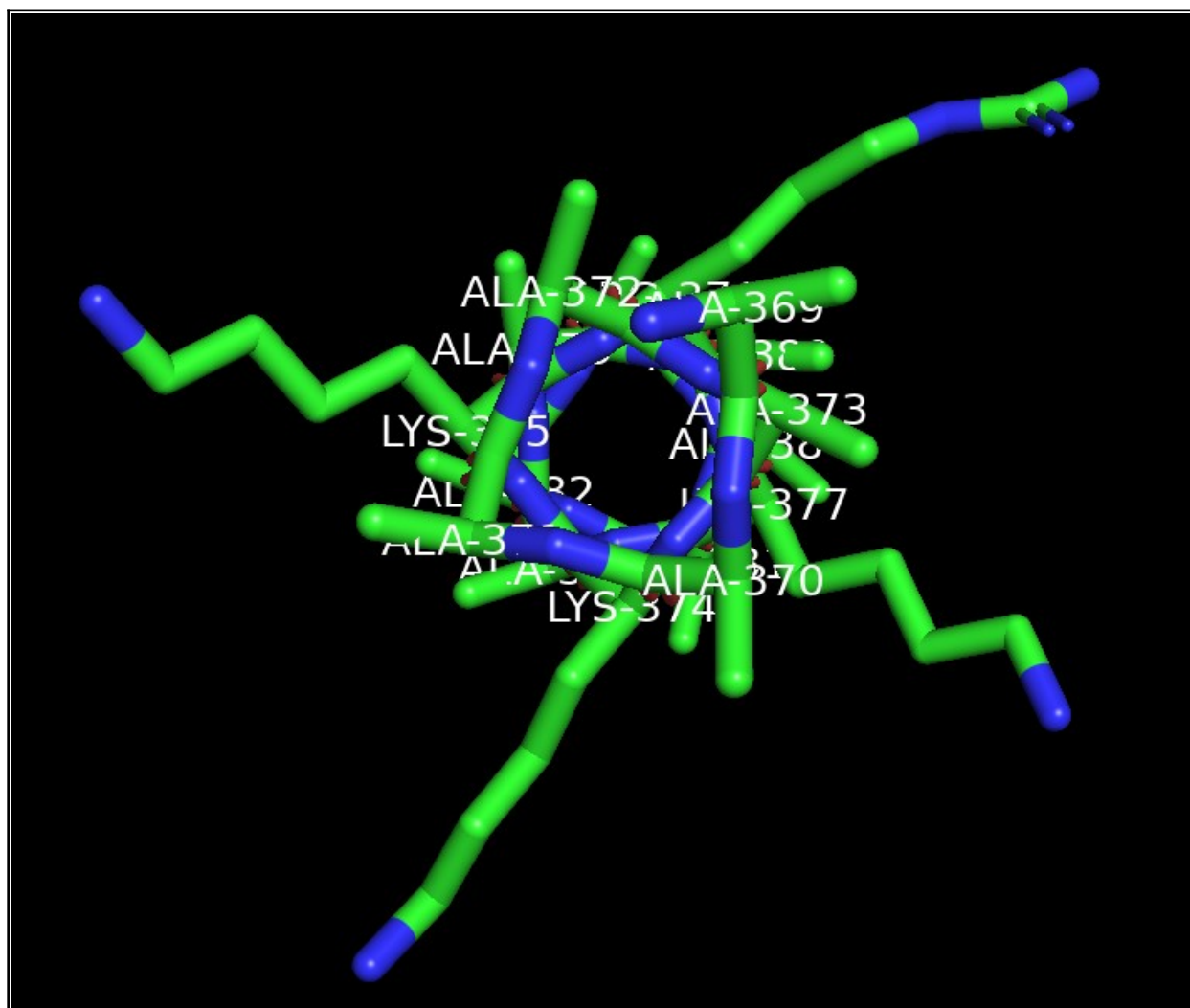


Figure 13d. Side view of amino acids seen in Figure 13c, showing heightened exposure of KKRK side chains when neighboring residues are mutated to alanines.

3.9) Distance heatmaps for residues flanking CALR WT KKRK

To further verify and quantify the difference in conformation when the neighboring residues around the KKRK are mutated to alanines, the same method was also applied to arginines as the positive charges between side chains were likely to repel each other and also result in a more open conformation. heatmaps were generated using the seaborn package in Python from the distances between the atoms at the end of each side chain measured by PyMOL. There were three heatmaps: for CALR WT as is (Figure 14a), for aspartic and glutamic acids mutated to arginines (Figure 14b), or to alanines (Figure 14c). Unsurprisingly, the distances were generally larger when the residues were mutated to arginines than alanines as a result of positively charged side chains repelling each other, although there were exceptions. The main focus of these heatmaps were the three glutamic and aspartic acids to either the left or right flank of the KKRK motif based on the approximate distance of 15 Å needed for possible interactions (refer to Table 2 for the full CALR sequence), with distances measured between the furthest atoms for each of the four amino acids inside the KKRK motif.

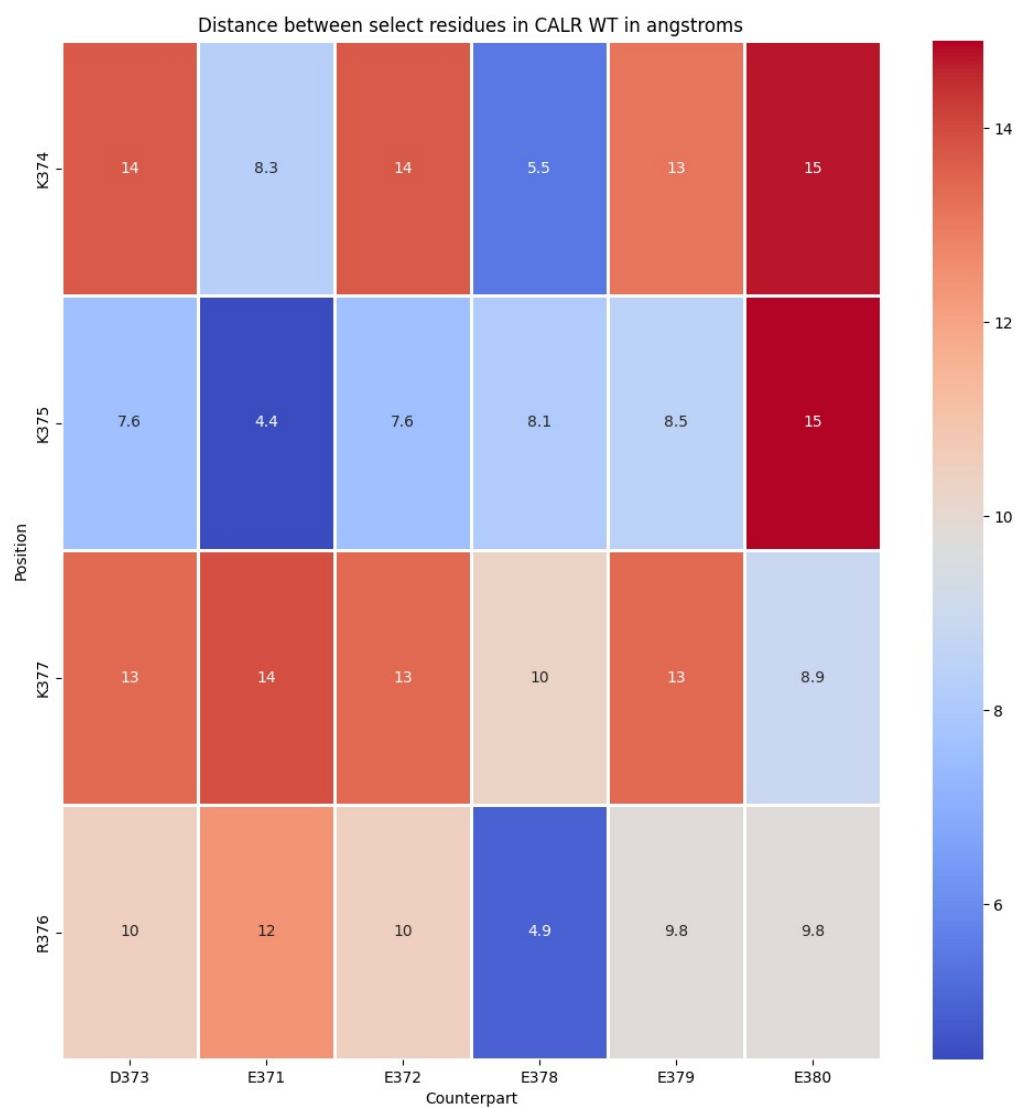


Figure 14a. Heatmap for CALR WT residues 371-380 without any mutations in PyMOL. Values are in angstroms.

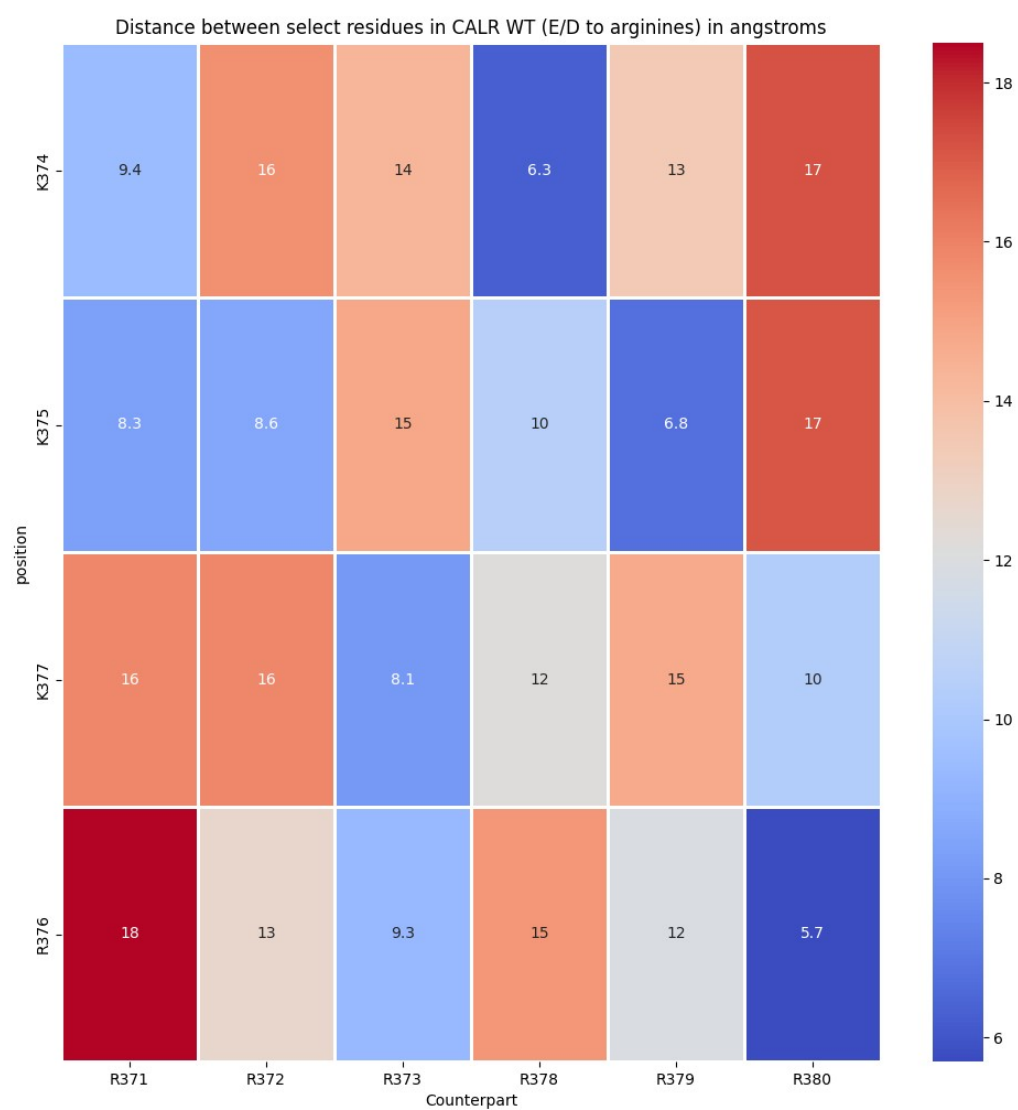


Figure 14b. Heatmap for CALR WT when aspartic acids and glutamic acids near KKRK are mutated to arginines.

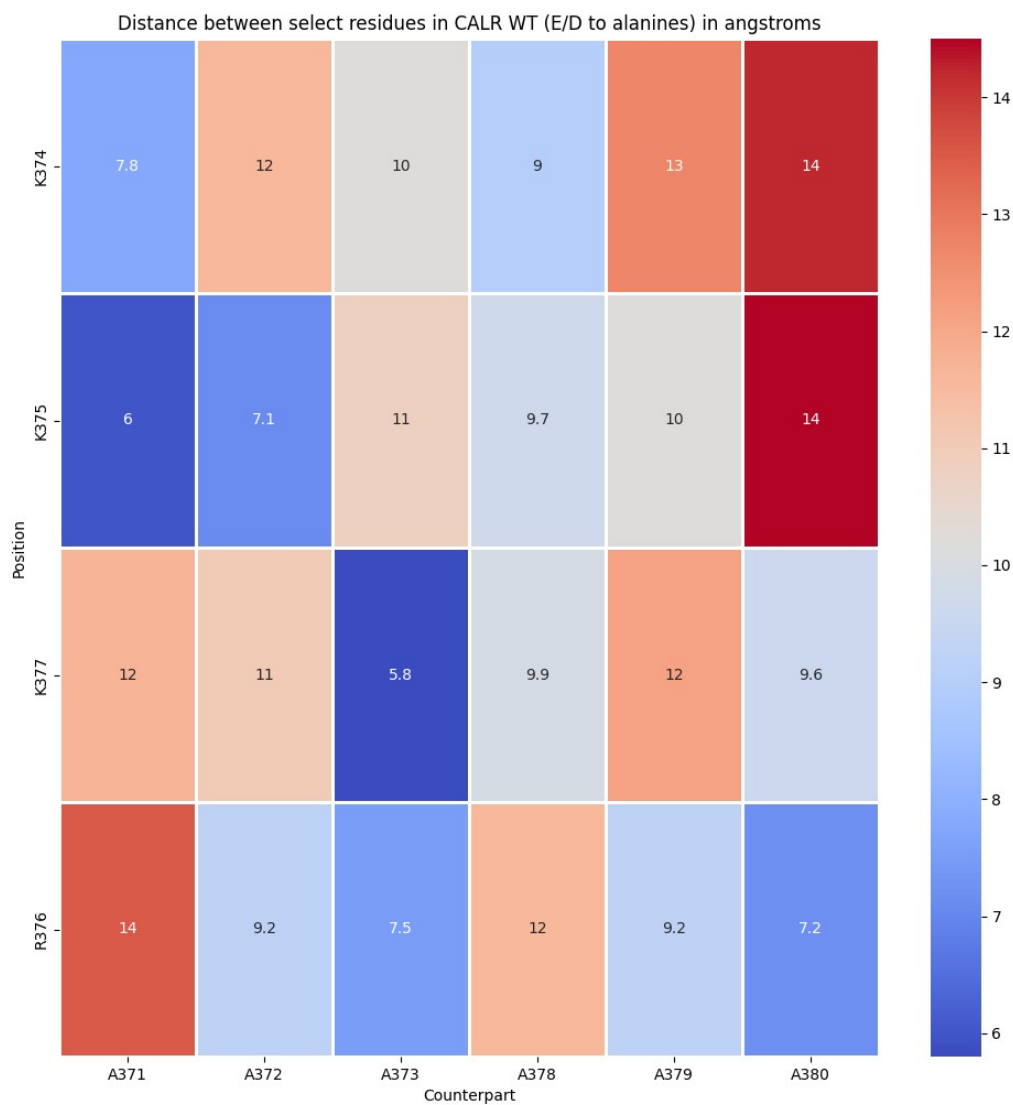


Figure 14c. Heatmap for CALR WT when aspartic acids and glutamic acids near KKRK are mutated to alanines.

3.10) Comparing effects of K/R tally versus KDEL tally on predicted CALR del52 secretion

Based on the interplay between newly acquired lysines and arginines and the KDEL sequence in terms of altering predicted sub-cellular localization as seen with the modified WT CALR (see Table 6 and Figures 9a-9b), a similar experiment was performed with CALR del52. Successive KDEL residues were added back onto the sequence at the very end of the C-terminus (ranging from one to seven KDEL motifs), while simultaneously using the K/R tally as before. Adding the KDEL motifs back into the sequence resulted in a decreased predicted secretion, but the effect of adding one motif versus seven was the same in terms of quantitative halting of predicted secretion (see Figures 15a-15b). Both MULocDeep and DeepLoc were employed for these analyses (see Figures 15a-15b), with the former showing a consistently higher range of values, similar to what was observed with each of the individual mutations for both CALR WT and del52. Due to this greater range of values, the multivariate correlation fit for the number of KDEL (henceforth referred to as KDEL tally) motifs and K/R tally was lower for MULocDeep than DeepLoc.

The KDEL tally did not make much of a difference, as secretion values generally remained similar as the number of KDEL residues increased as the K/R tally remained the same. K/R tally was shown to be a stronger contributing factor than the KDEL tally based on the higher value of its t-value (see Figures 15c and 15d) in both DeepLoc and MULocDeep analyses. Similar to what was observed when comparing individual secretion motifs' effects on secretion, the drop in secretion was more consistent and conservative in DeepLoc than MULocDeep, hence the higher t-value for the former over the latter. However, both showed that even in the presence of several KDEL anchor residues, secretion still occurred, presumably because of these arginine-rich motifs.

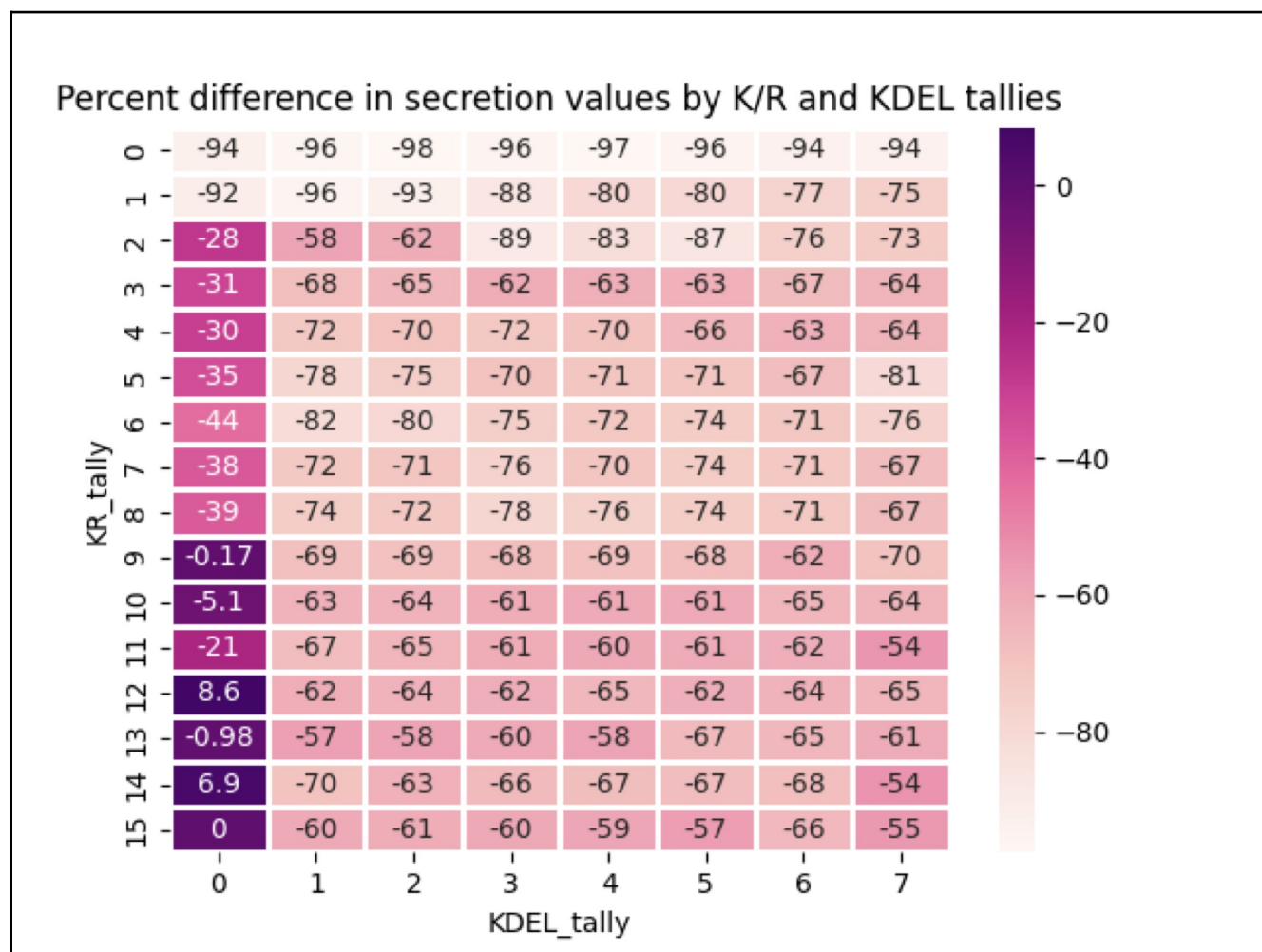


Figure 15a. MULocDeep analysis of secretion scores normalized to CALR del52 factoring in K/R tally and KDEL tally. Values inside heatmap are percent differences in predicted secretion compared to CALR del52.

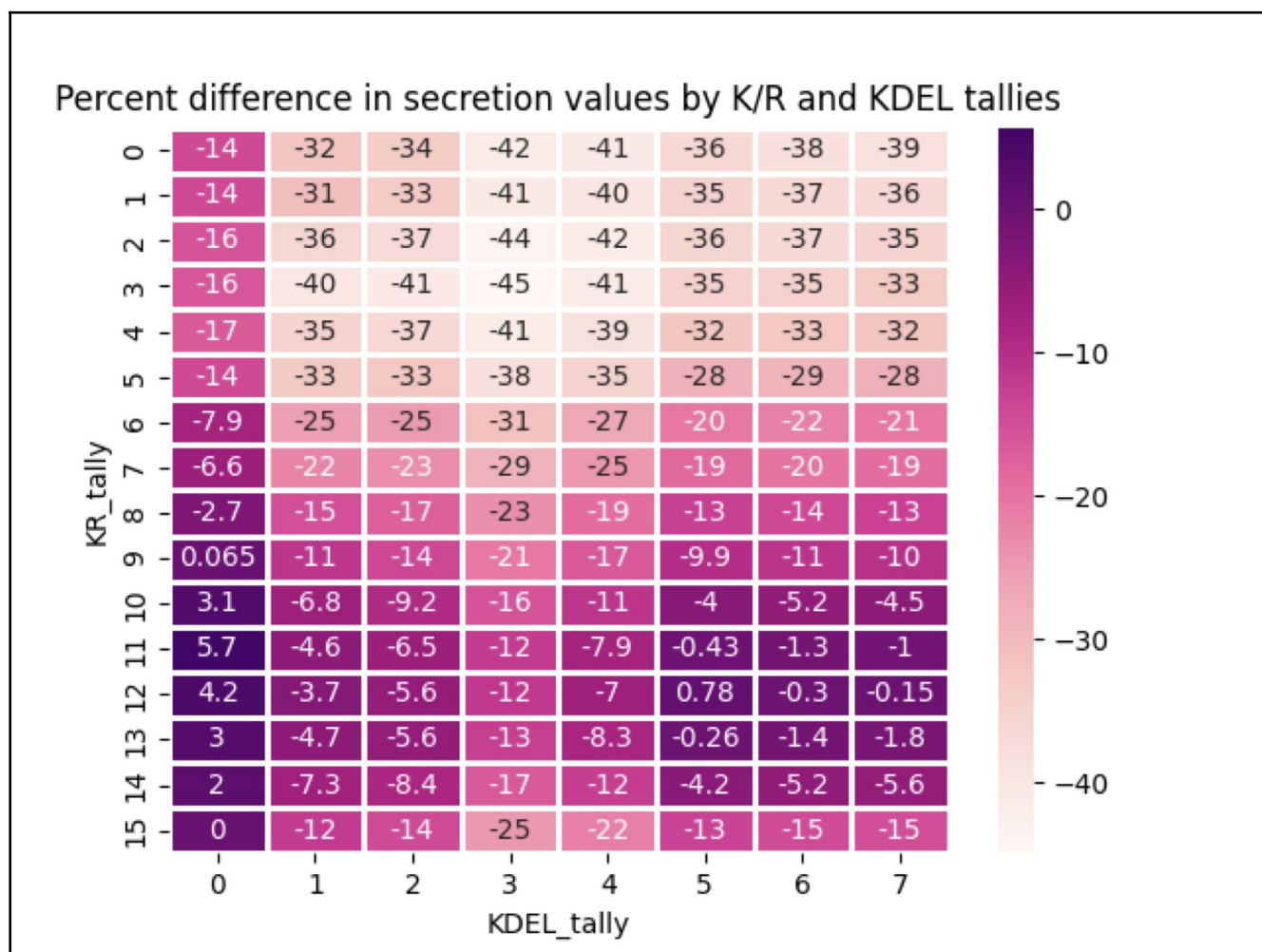


Figure 15b. DeepLoc analysis of secretion scores normalized to CALR del52 factoring in K/R tally and KDEL tally. Values inside heatmap are percent differences in predicted secretion compared to CALR del52.

OLS Regression Results						
Dep. Variable:	secretion_diff	R-squared:	0.659			
Model:	OLS	Adj. R-squared:	0.653			
Method:	Least Squares	F-statistic:	120.5			
Date:	Sun	15 Oct 2023	Prob (F-statistic):	6.79e-30		
Time:	09:57:48	Log-Likelihood:	-449.83			
No. Observations:	128	AIC:	905.7			
Df Residuals:	125	BIC:	914.2			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	-34.5592	1.778	-19.437	0.000	-38.078	-31.040
KDEL_tally	-0.8312	0.317	-2.619	0.010	-1.459	-0.203
KR_tally	2.4139	0.158	15.305	0.000	2.102	2.726
Omnibus:	0.328	Durbin-Watson:	0.551			
Prob(Omnibus):	0.849	Jarque-Bera (JB):	0.427			
Skew:	-0.114	Prob(JB):	0.808			
Kurtosis:	2.831	Cond. No.	23.1			
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Figure 15c. OLS regression analysis results for data seen in Figure 15b.

OLS Regression Results						
Dep. Variable:	secretion_diff	R-squared:	0.359			
Model:	OLS	Adj. R-squared:	0.348			
Method:	Least Squares	F-statistic:	34.93			
Date:	Thu	19 Oct 2023	Prob (F-statistic):	8.87e-13		
Time:	11:55:28	Log-Likelihood:	-537.19			
No. Observations:	128	AIC:	1080.			
Df Residuals:	125	BIC:	1089.			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	-69.1821	3.518	-19.664	0.000	-76.145	-62.219
KDEL_tally	-3.1980	0.628	-5.093	0.000	-4.441	-1.955
KR_tally	2.0684	0.312	6.628	0.000	1.451	2.686
Omnibus:	26.222	Durbin-Watson:	0.461			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	35.696			
Skew:	1.100	Prob(JB):	1.77e-08			
Kurtosis:	4.360	Cond. No.	23.1			
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Figure 15d. OLS regression analysis results for data seen in Figure 15a.

3.11) Comparison of CALR ins5 and CALR ins5_alanine

The next analysis compared the other main mutant of CALR, ins5 (see Table 2), with its negatively charged residues (aspartic and glutamic acids) mutated to alanines, henceforth referred to as ins5_alanine, similar to what was done for the secretion motifs in del52. Upon these mutations, the predicted secretion levels of ins5 increased compared to del52: in MULocDeep, ins5_alanine scored better compared to ins5 but still lower than del52. Recent studies have shown that even with higher allele burden, ins5 has lower levels of secretion than del52, corroborating the MULocDeep prediction [90]. Similar to when CALR WT lost its aspartic and glutamic acids around the KKRR motif, a similar effect occurred with CALR ins5, which has its frameshift mutation further downstream than del52 (see Table 2 for full sequences) and thus retains these residues. In the DeepLoc sorting signal graph for ins5_alanine, the KKRR motif had a similar score compared to the motifs shared between ins5 and del52 (Figure 16a), further suggesting that it could also function in a similar way to our proposed motifs. This has also been shown in the modified CALR WT sequence ELM search results (Table 9b). However, the modest increase in predicted secretion for ins5_alanine also suggests that adding additional motifs or removing acidic residues is not always effective as a strategy for improving secretion, in line with the results from Figures 11a-11b.

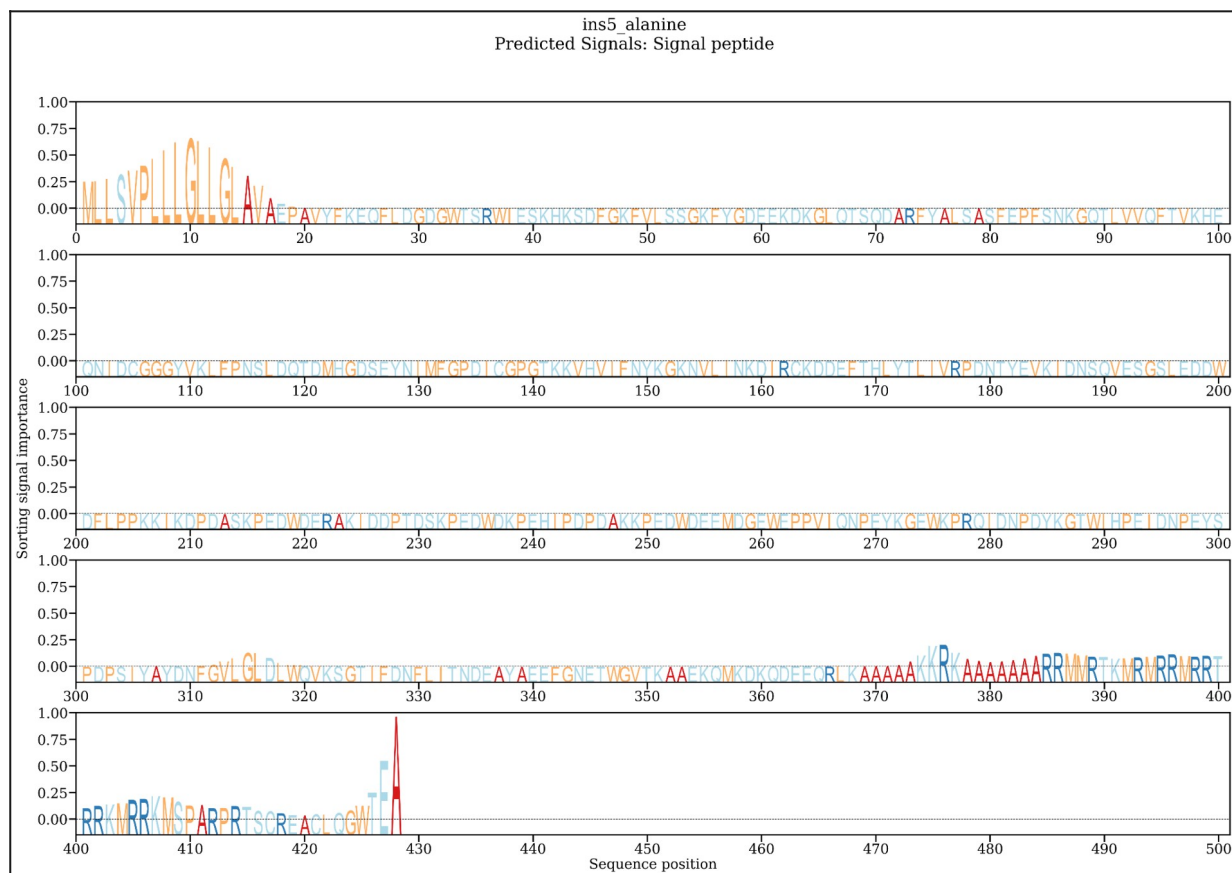


Figure 16a. DeepLoc sorting signal graph for CALR ins5_alanine.

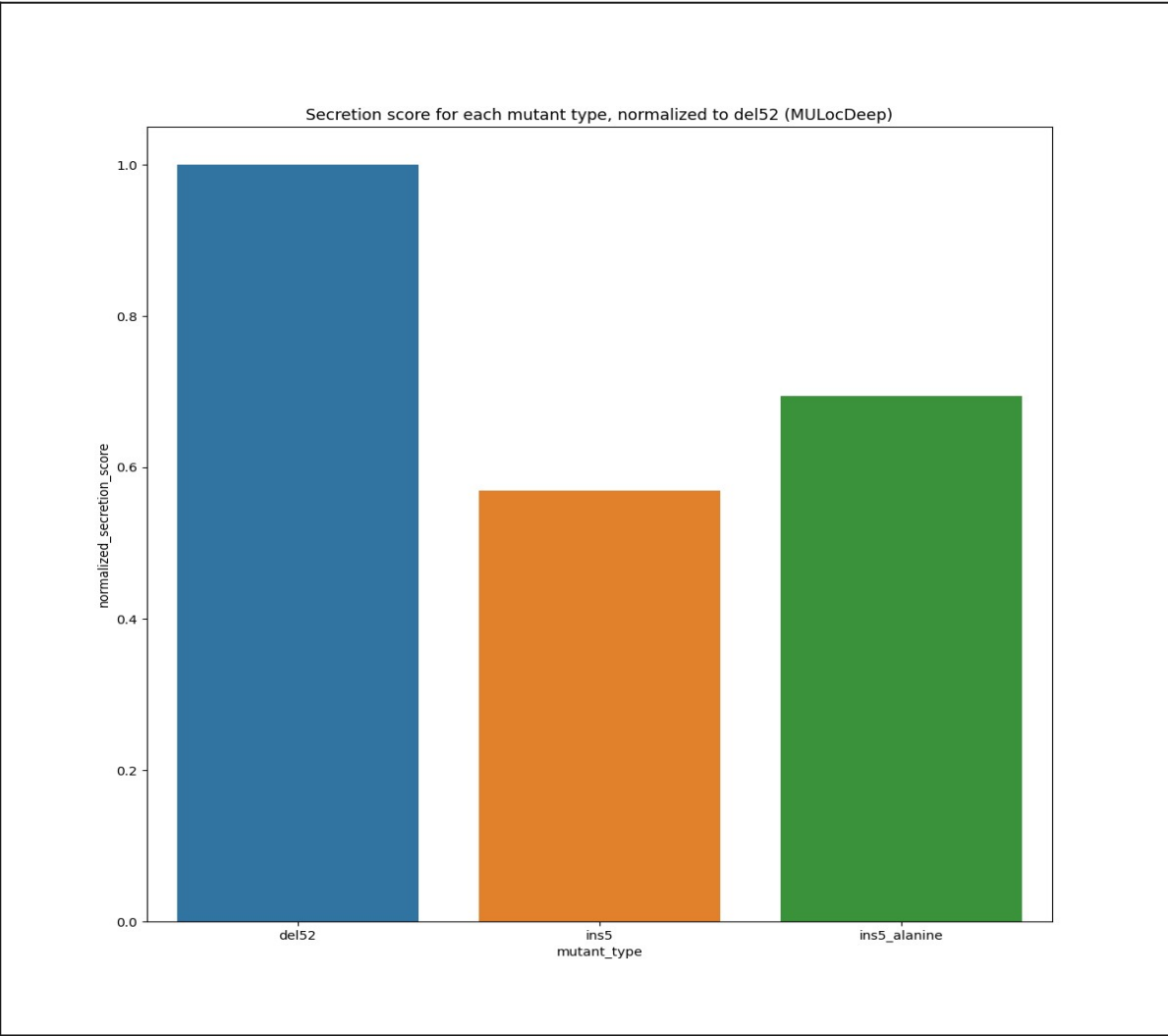


Figure 16b. Predicted secretion scores for CALR ins5, CALR del52, and CALR ins5_alanine according to MULocDeep.

3.12) Comparison of lysines and arginines' ability to affect CALR del52 secretion

There is an underlying assumption about the K/R tally analyses and all other *in silico* substitutions performed up to this point, and that is the functional similarity of lysines and arginines. However, a study by Li et al. found that arginines are more effective at traversing the cell membrane than lysines [121]. Indeed, there are twelve arginines and three lysines in the mutant CALR C-terminus, which may partially explain how del52 and ins5 gain the ability to be secreted. To check for the role of each amino acid type in secretion, arginines in the del52 sequence were mutated to lysines and vice versa. DeepLoc was used for this analysis due to its cleaner and more consistent trends compared to MULocDeep. For the prediction when arginines were mutated to lysines (labeled as extra_K, see Figure 17a), the loss of arginines resulted in a similar change in predicted secretion compared to that seen in Figures 11c-11d when arginines (and lysines) were mutated to alanines, further suggesting that the means of loss of arginines does not matter: the end result is similar whether they are mutated to lysines, methionines, or alanines. When lysines were mutated to arginines (see Figure 17b), there was only a modest increase in predicted secretion (~ 9%), since three residues were mutated in total. These results confirm the key role of arginines in secretion, with lysines possibly playing a secondary role that may facilitate the process, but this needs to be investigated in further detail.

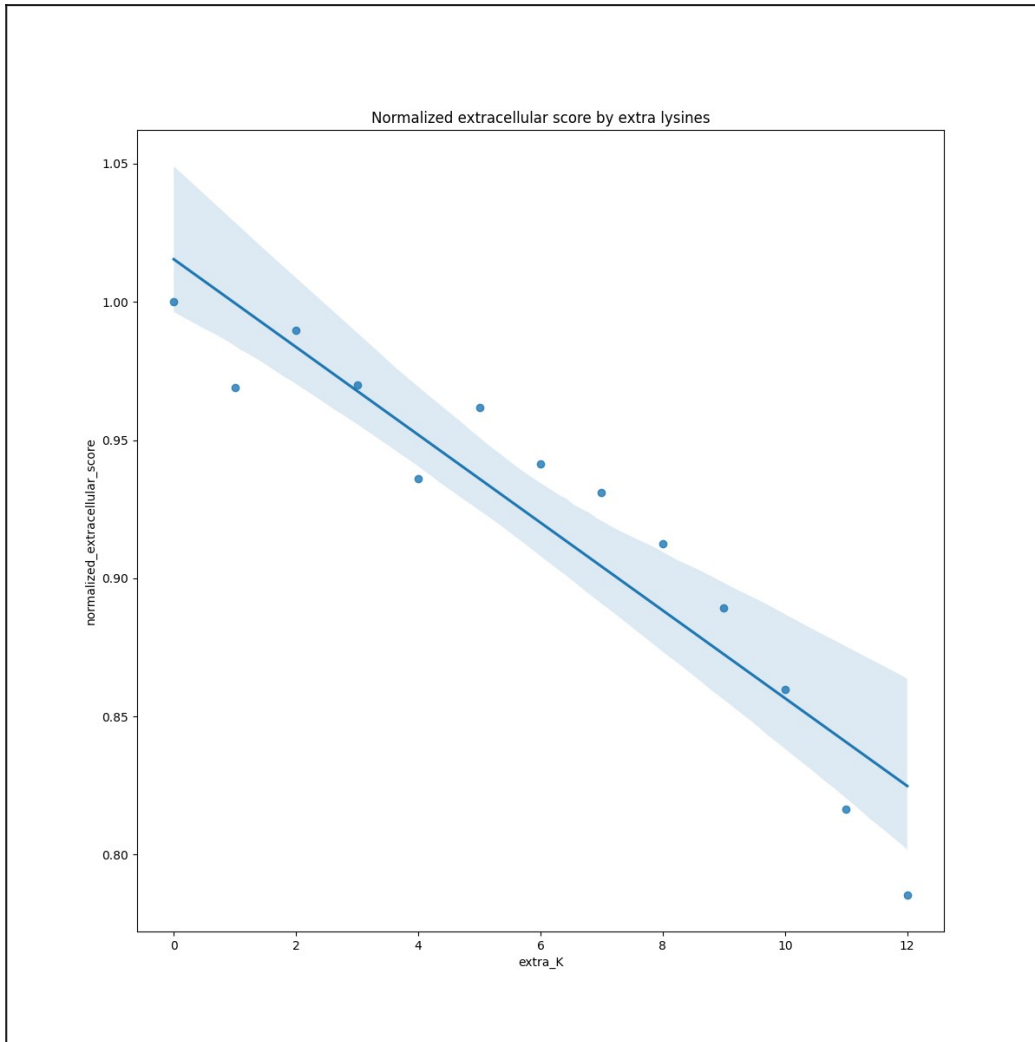


Figure 17a. DeepLoc prediction of extracellular localization for CALR del52 according to the number of arginines substituted with lysines (“extra_K”). Scores normalized to CALR del52 (extra_K = 0).

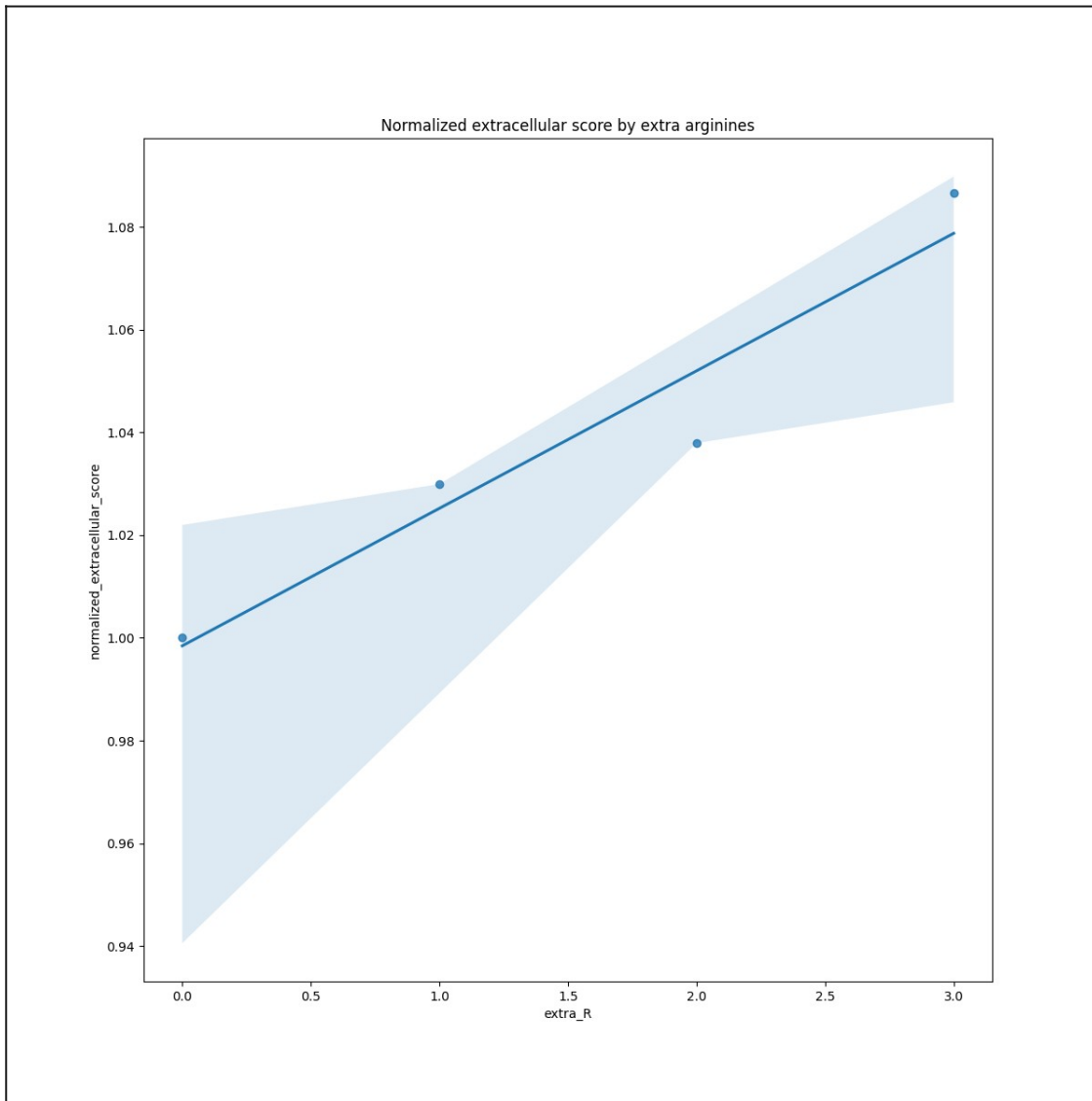


Figure 17b. DeepLoc prediction of extracellular localization based on the number of lysines mutated to arginines. Scores normalized to CALR del52 (extra_R = 0).

3.13) Investigation of a possible correlation between PTM and K/R tally

In addition to predicting levels of secretion based on either the motifs or K/R tally, further analyses were employed to screen for effects on secretion by PTM: in particular the serines at positions 392 and 399 and the threonine at position 373 (see Table 2 for full del52 sequence). These residues were selected because they are only present in either type of CALR mutant but not in WT CALR. Based on the results so far, altering the overall charge of the C-terminus is thus sufficient to affect the ability of mutant CALR to move outside of the ER. However, two questions that arose were whether or not the presence or absence of these PTM also had an effect on predicted secretion, and also whether the K/R tally affected the probability of PTM at these residues. Two types of PTM were examined using the web-based MuSiteDeep predictor [112], and later on DeepPhos [116]: O-linked glycosylation and phosphorylation. For all three residues, there was a strong inverse correlation between glycosylation and K/R tally (see Figures 18a-18f). Additionally, all three residues were more likely to be phosphorylated when the K/R tally was higher (see Table 10 for correlation values). DeepPhos predictions (Figures 19a-19b) showed a similar pattern for S392 and S399, but for T373 the correlation was much weaker (~ 0.4 , see Table 11 and Figure 19c). Due to both having a strong propensity for being phosphorylated, the correlation between S392 and S399 was very high (~ 0.9 , see Tables 11 and 12), suggesting that both serines are likely phosphorylated in tandem.

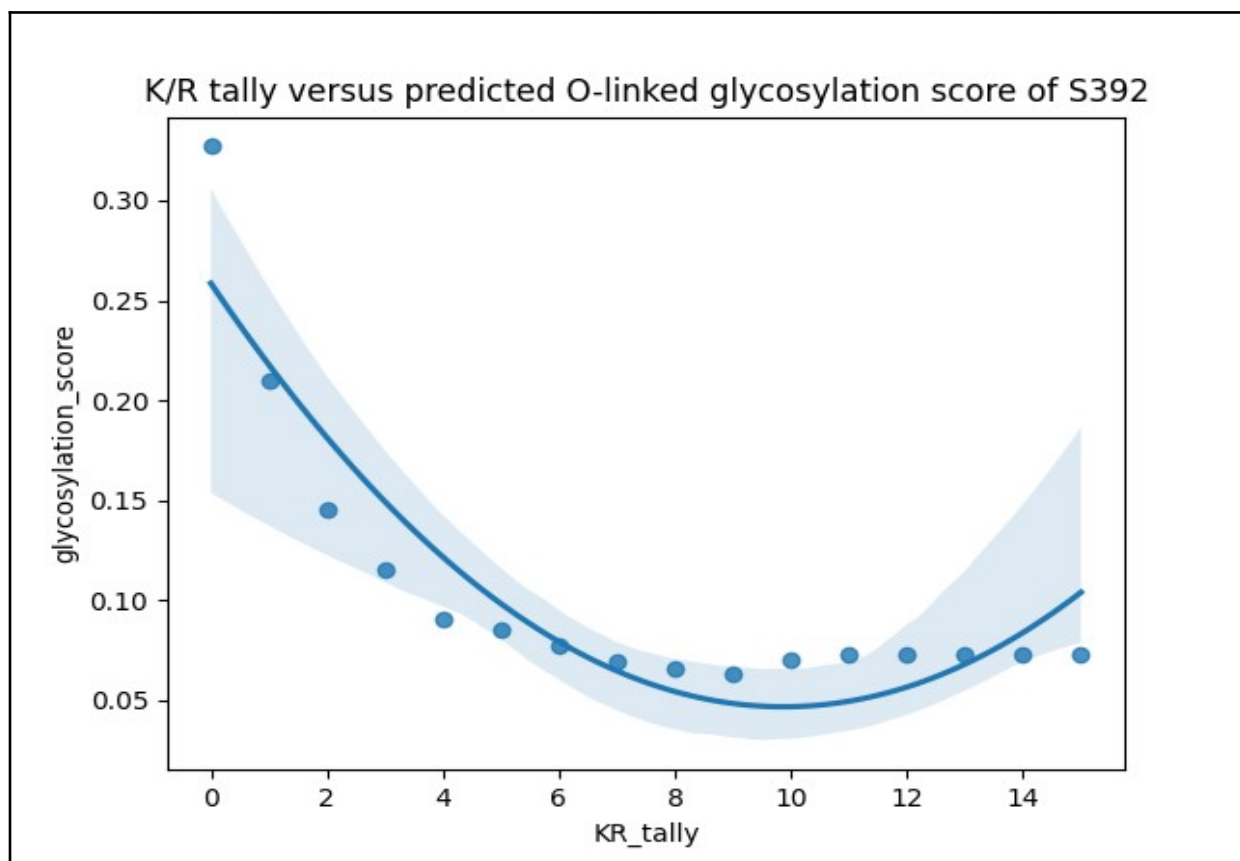


Figure 18a. Predicted O-linked glycosylation score for S392 based on K/R tally according to MuSiteDeep.

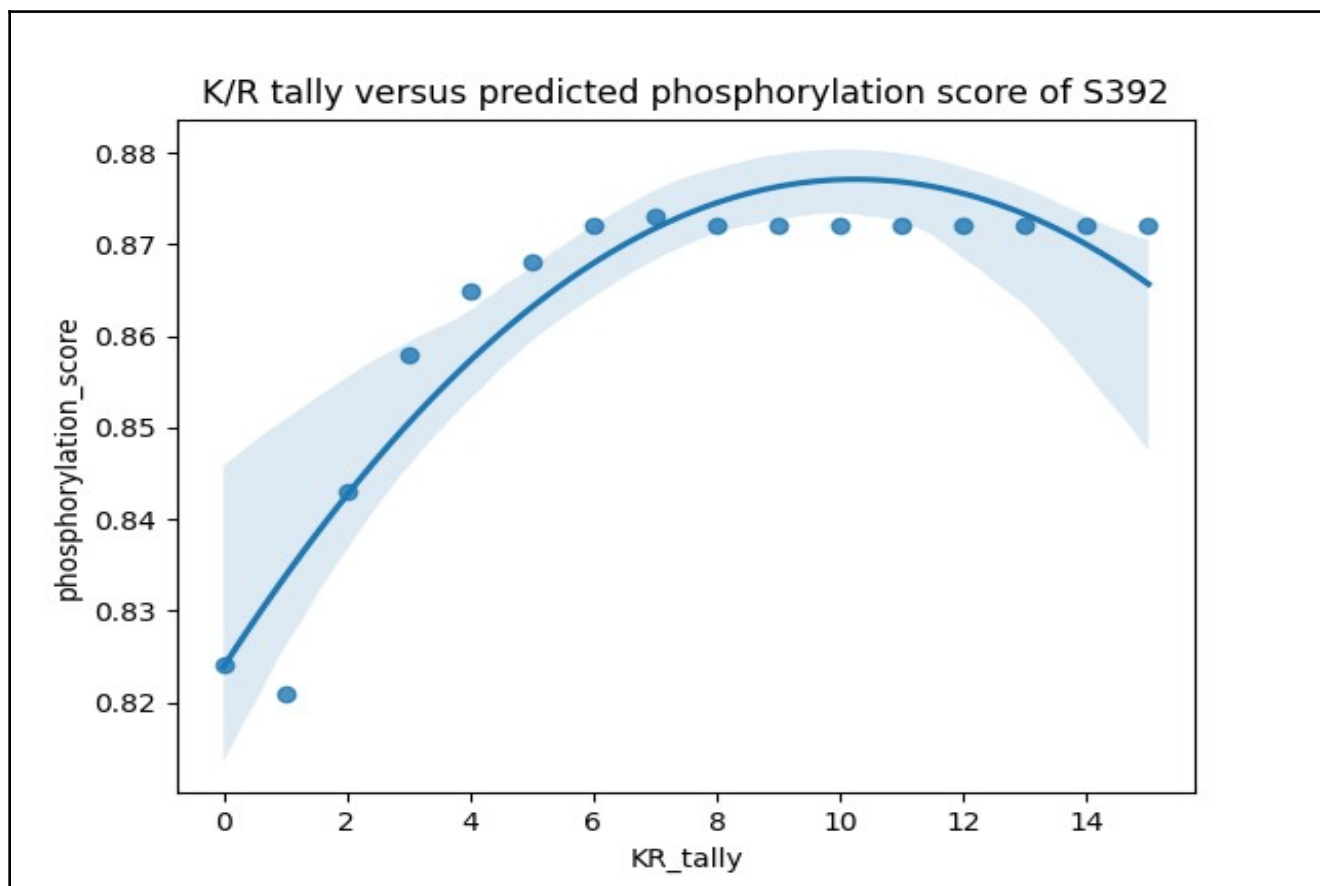


Figure 18b. Predicted phosphorylation score for S392 based on K/R tally according to MuSiteDeep.

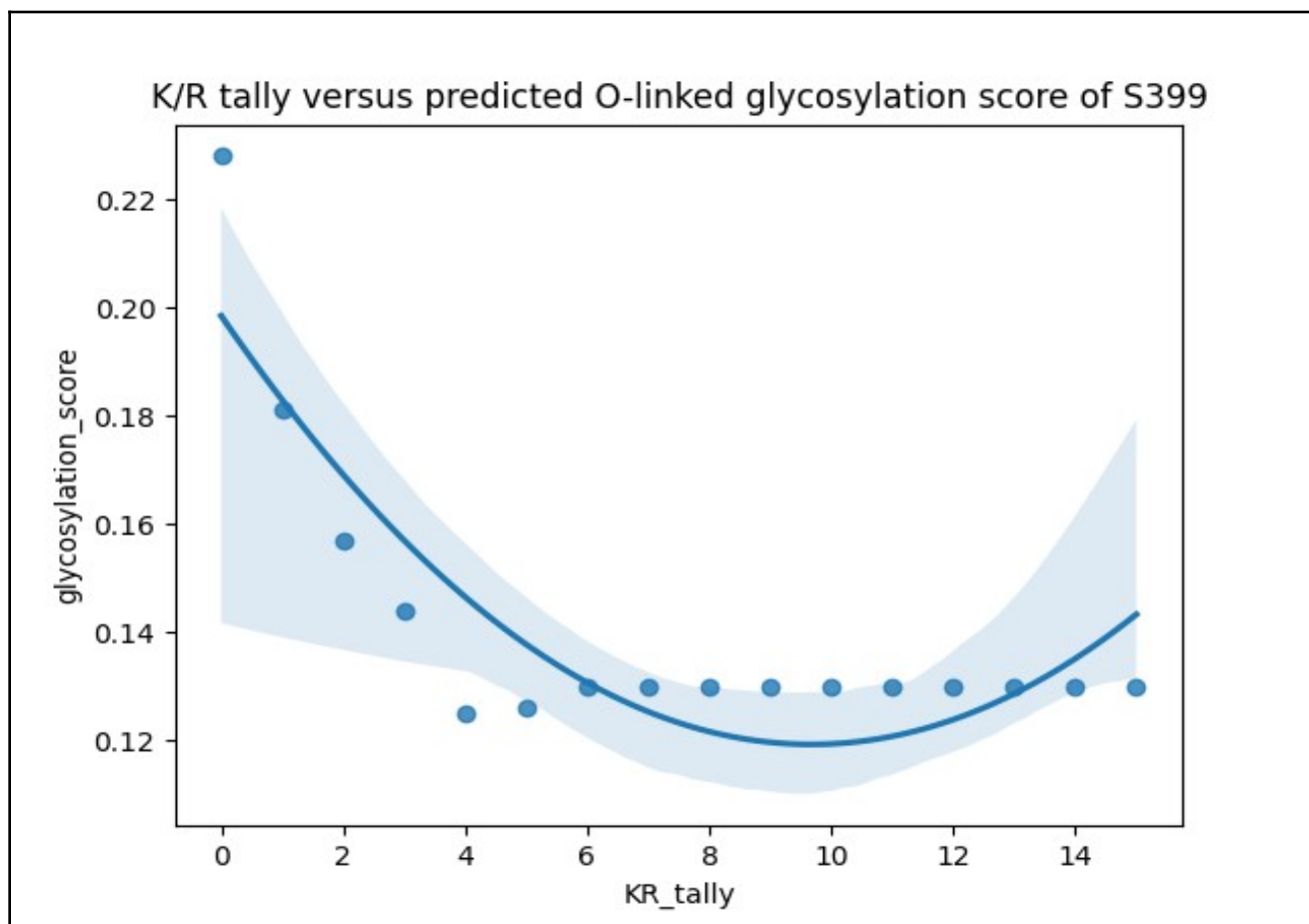


Figure 18c. Predicted O-linked glycosylation for S399 based on K/R tally according to MuSiteDeep.

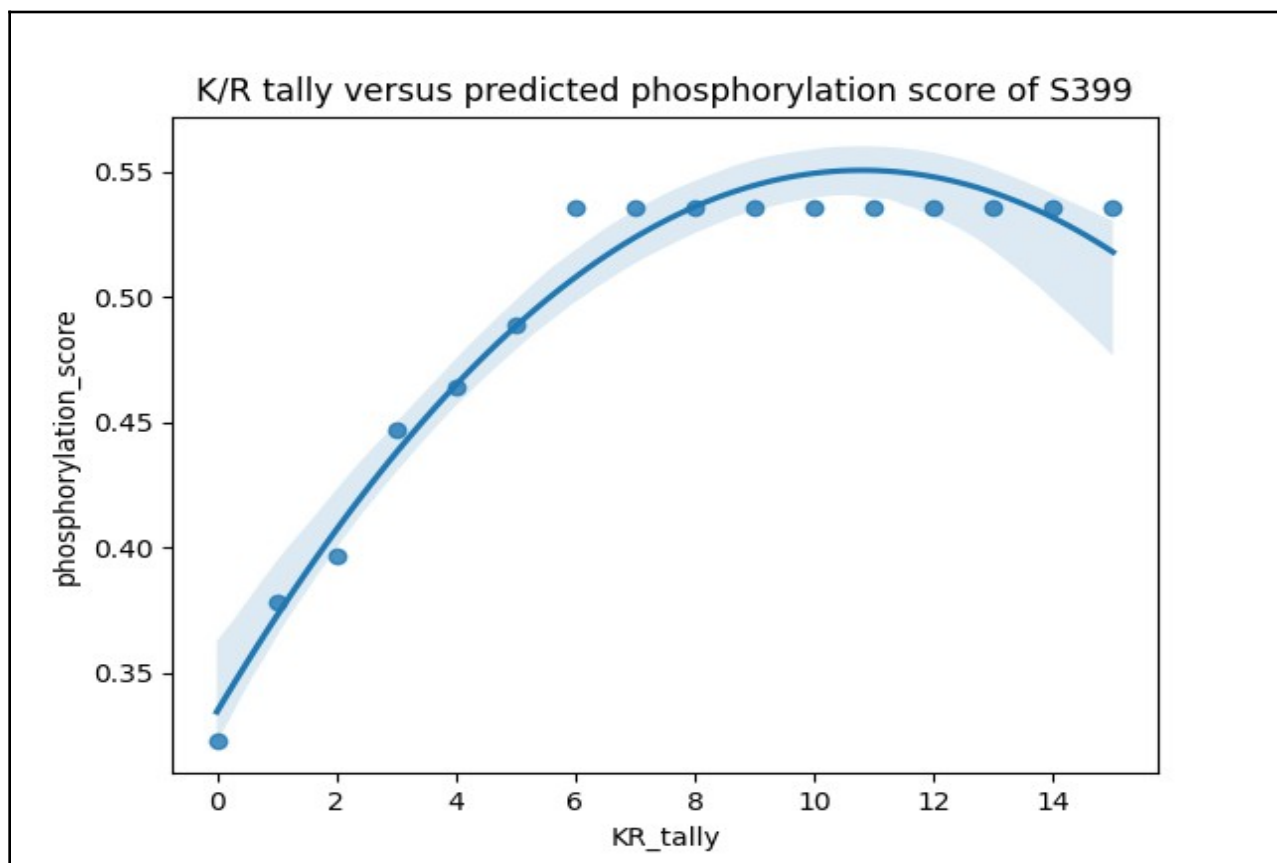


Figure 18d. Predicted phosphorylation for S399 based on K/R tally according to MuSiteDeep.

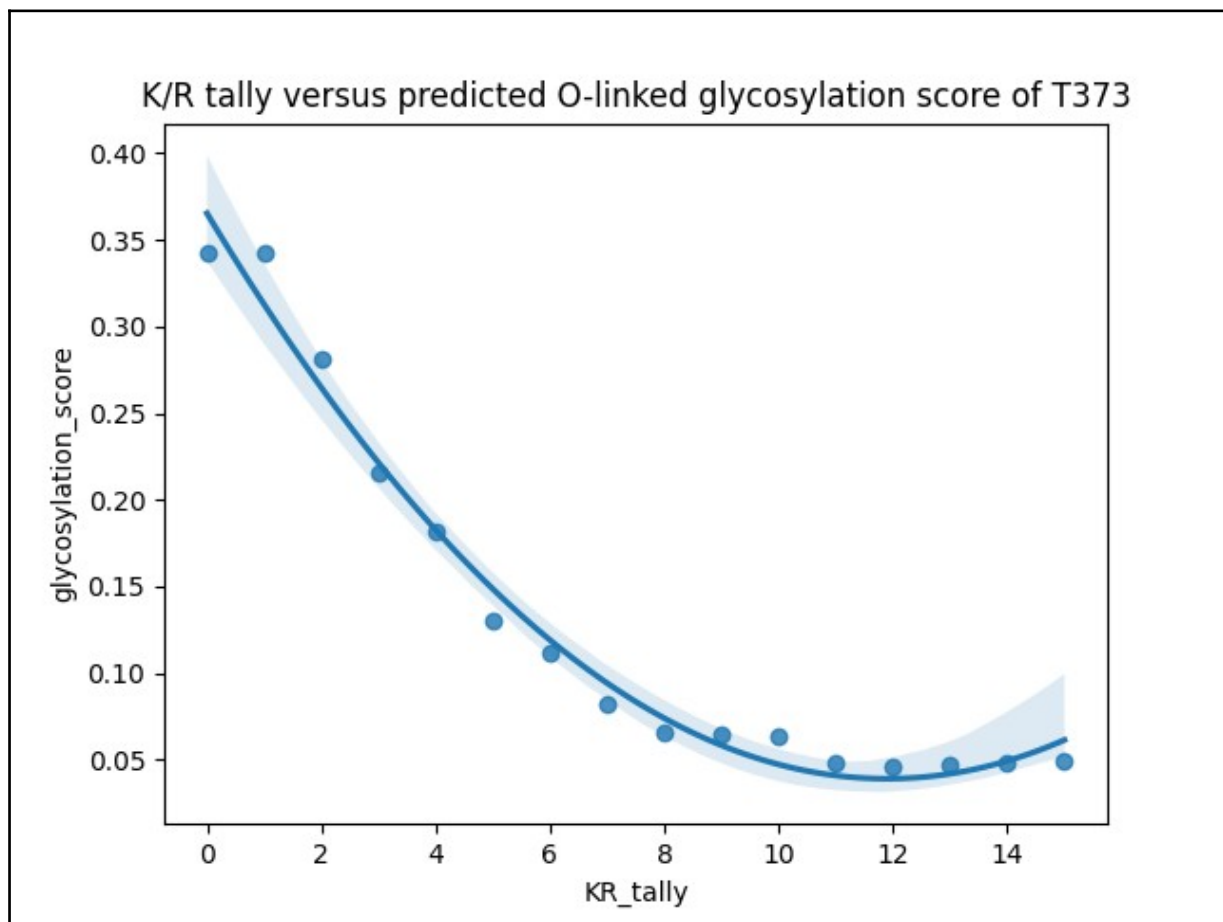


Figure 18e. Predicted O-linked glycosylation for T373 based on K/R tally according to MuSiteDeep.

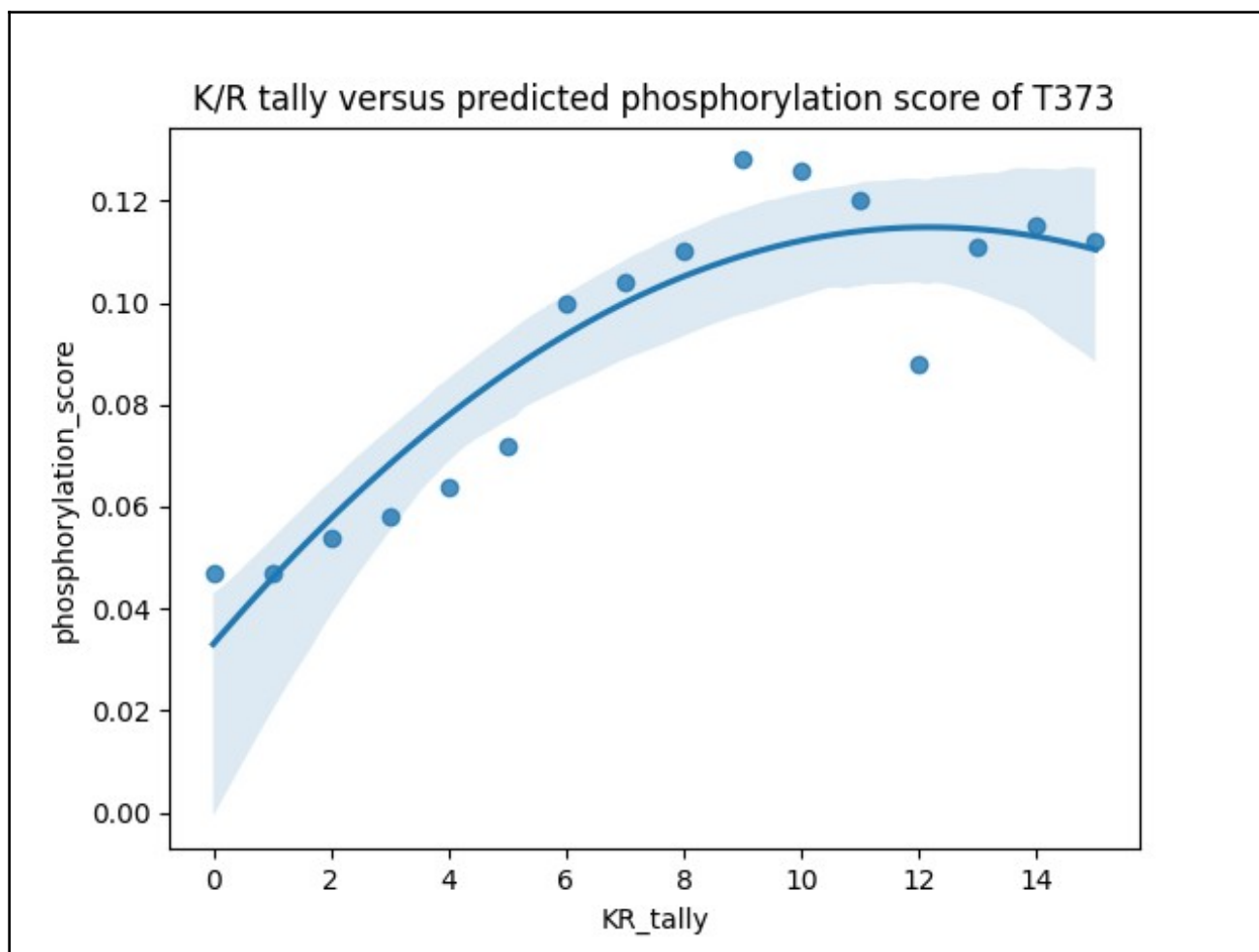


Figure 18f. Predicted phosphorylation for T373 based on K/R tally according to MuSiteDeep.

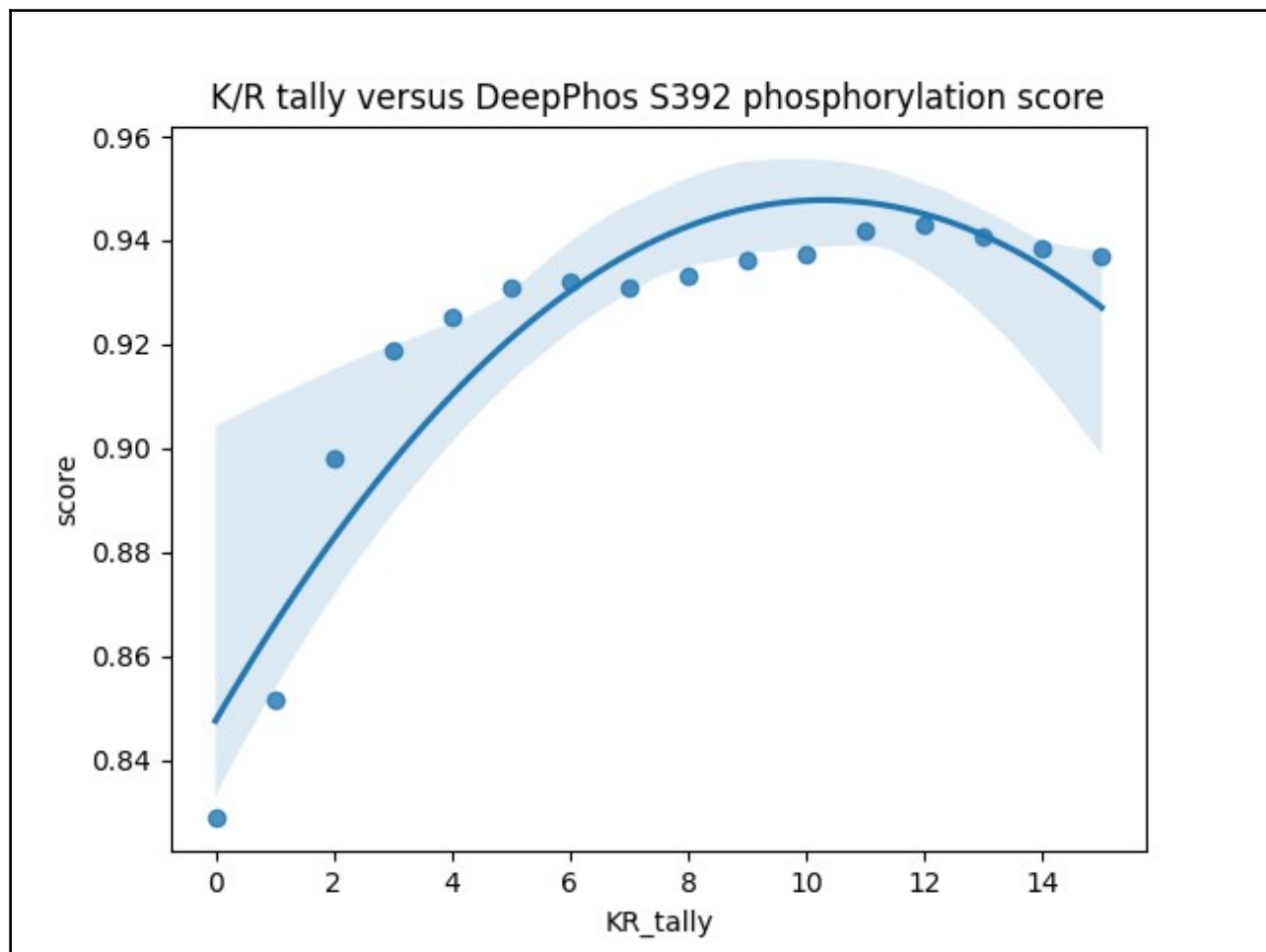


Figure 19a. DeepPhos phosphorylation predictions for S392 based on K/R tally.

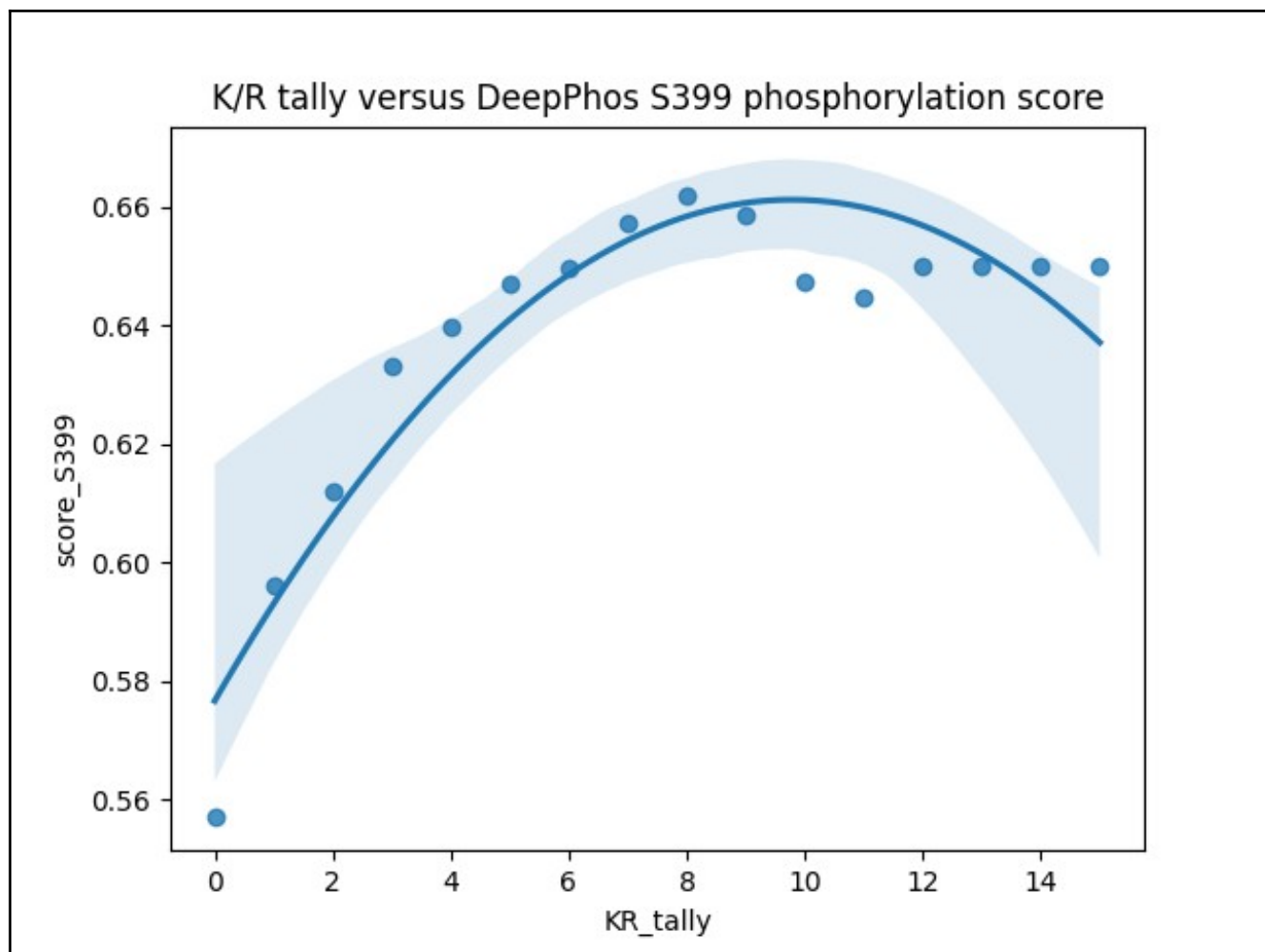


Figure 19b. DeepPhos phosphorylation predictions for S399 based on K/R tally.

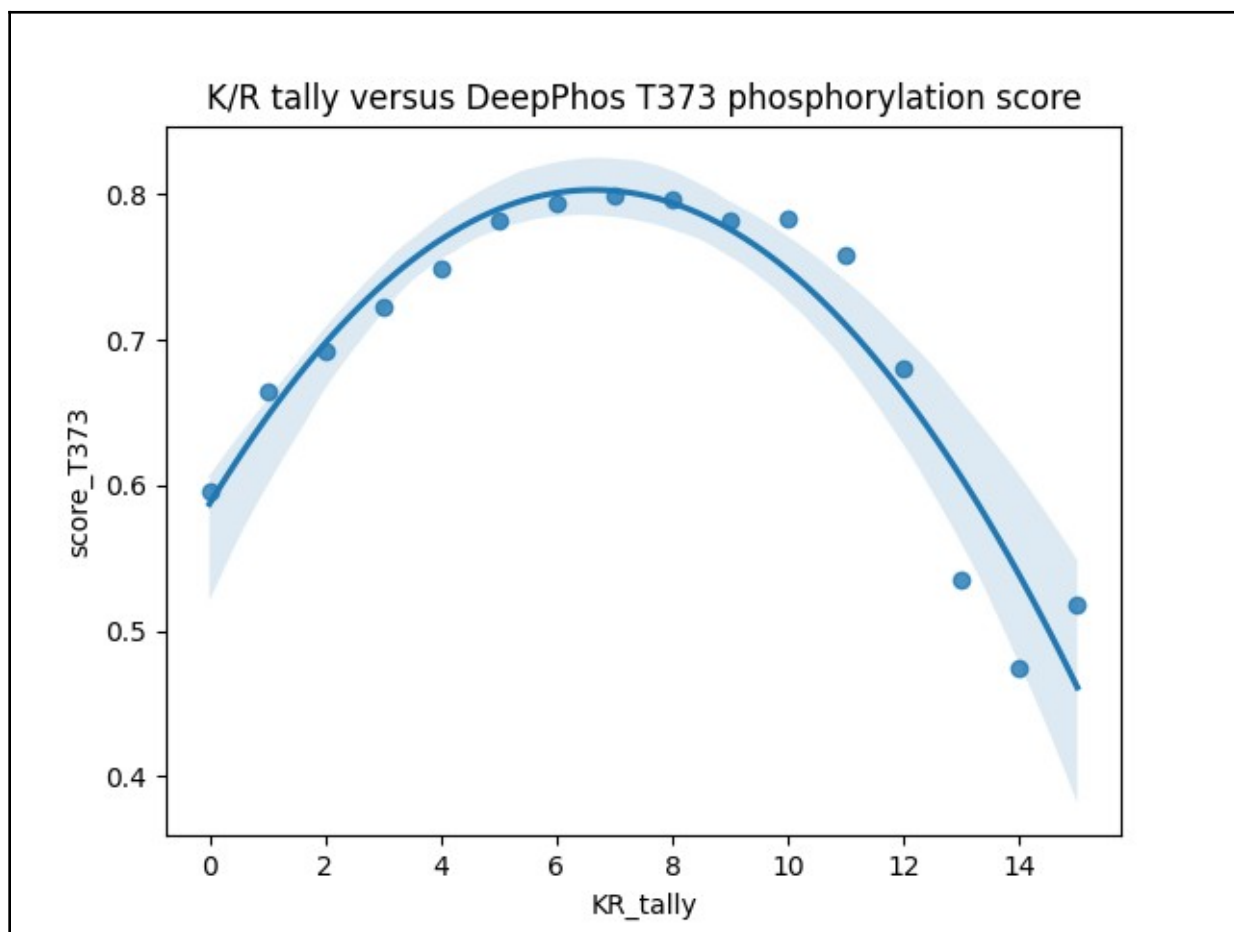


Figure 19c. DeepPhos phosphorylation predictions for T373 based on K/R tally.

Table 10. Pearson correlation values for MuSiteDeep data seen in Figures 18a-f.

	KR_tally	phosphorylation_score_S392	glycosylation_score_S392	phosphorylation_score_S399	glycosylation_score_S399	phosphorylation_score_T373	glycosylation_score_T373
KR_tally	1.0	0.7610199834217850	-0.6969892276106360	0.8362568322885030	-0.6430203656698880	0.8377799878181250	-0.8985001967244370
phosphorylation_score_S392	0.7610199834217850	1.0	-0.9307093376137220	0.9565656609710570	-0.9141597384809550	0.8042328416009650	-0.9501372920534090
glycosylation_score_S392	-0.6969892276106360	-0.9307093376137220	1.0	-0.9281413394670040	0.9894214469778710	-0.7298908738534080	0.8715663264654690
phosphorylation_score_S399	0.8362568322885030	0.9565656609710570	-0.9281413394670040	1.0	-0.8880037491778090	0.889514670072998	-0.9769049379931240
glycosylation_score_S399	-0.6430203656698880	-0.9141597384809550	0.9894214469778710	-0.8880037491778090	1.0	-0.6518518299281010	0.8275371505277030
phosphorylation_score_T373	0.8377799878181250	0.8042328416009650	-0.7298908738534080	0.889514670072998	-0.6518518299281010	1.0	-0.9060466571826410
glycosylation_score_T373	-0.8985001967244370	-0.9501372920534090	0.8715663264654690	-0.9769049379931240	0.8275371505277030	-0.9060466571826410	1.0

Table 11. Pearson correlation values for data seen in Figures 19a-c.

	KR_tally	score_S399	score_T373	score_S392
KR_tally	1.0	0.6976455628162770	-0.3671933815918720	0.7564825262368830
score_S399	0.6976455628162770	1.0	0.26372205854342500	0.9605381021023320
score_T373	-0.3671933815918720	0.26372205854342500	1.0	0.17580426833577100
score_S392	0.7564825262368830	0.9605381021023320	0.17580426833577100	1.0

3.14) Examination of the loss of potential PTM on CALR del52 secretion

The second question of these PTMs being a prerequisite for mutant CALR secretion was answered using MULocDeep and DeepLoc with modified versions of del52: S392A, T373A, and S399A plus combinations of these mutations while still using K/R tally as a gradient. None of these resulted in a significant difference in secretion according to either MULocDeep or DeepLoc (see Figures 20a-20b). Based on t-values, which is a marker for statistical significance, the K/R tally still had a much higher impact (see Figures 20c and 20d) than any of the mutagenesis types, suggesting that the PTMs are a consequence of secretion, rather than a prerequisite. This result was shared between both DeepLoc and MULocDeep, but more pronounced for DeepLoc due to the more conservative, and thus more consistent trends in secretion drops.

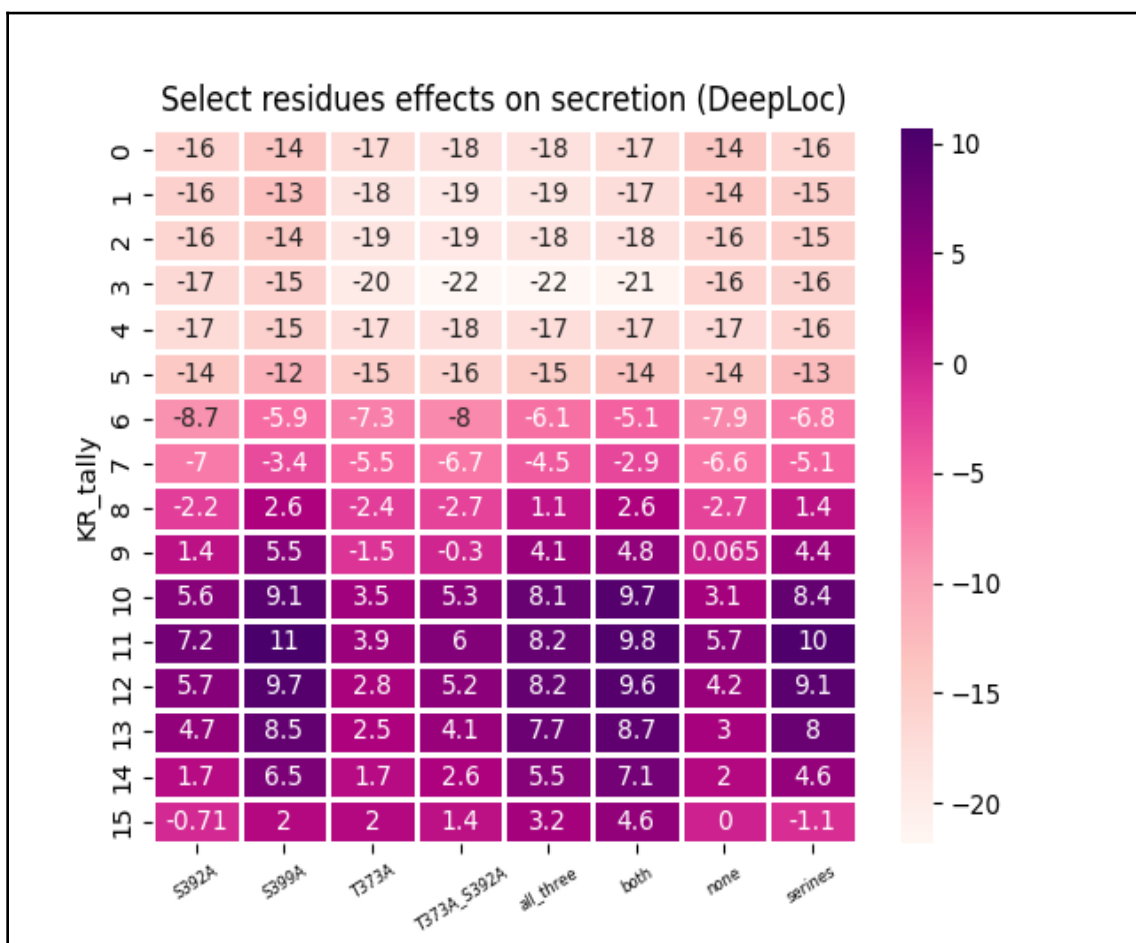


Figure 20a. Heatmap showing percentage difference compared to CALR del52 using DeepLoc predictions. “None” refers to none of these residues being mutated to alanines, “serines” referring to S392 and S399, and “both” refers to S399A and T373A present in the sequence.

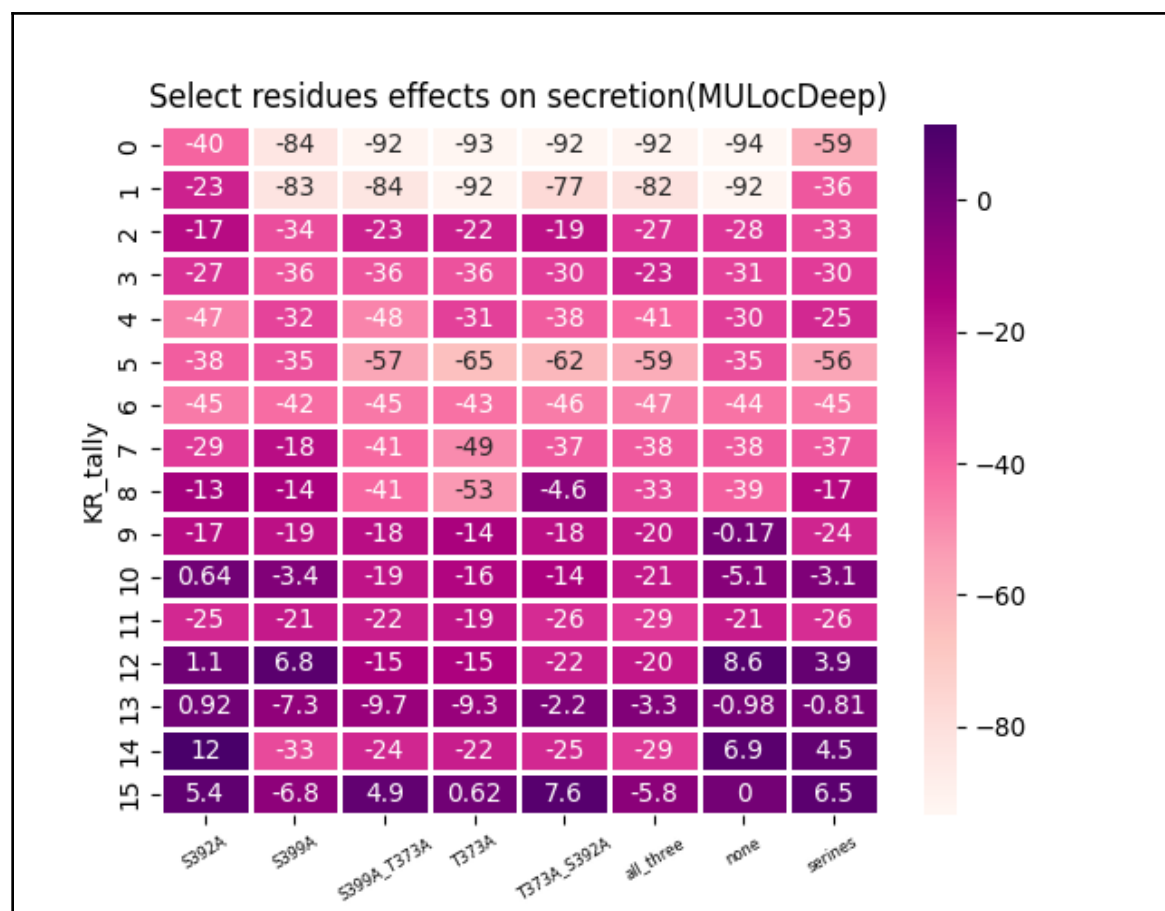


Figure 20b. Heatmap showing percentage difference compared to CALR del52 using MULocDeep predictions. “None” refers to none of these residues being mutated to alanines, “serines” refer to S392 and S399, and “both” refers to S399A and T373A present in the sequence.

OLS Regression Results						
Dep. Variable:	normalized_extracellular_score	R-squared:	0.807			
Model:	OLS	Adj. R-squared:	0.794			
Method:	Least Squares	F-statistic:	62.01			
Date:	Wed	15 May 2024	Prob (F-statistic):	7.34e-39		
Time:	11:42:28	Log-Likelihood:	217.30			
No. Observations:	128	AIC:	-416.6			
Df Residuals:	119	BIC:	-390.9			
Df Model:	8					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.7998	0.013	60.346	0.000	0.774	0.826
mutagenesis_type[T.S399A]	0.0307	0.016	1.888	0.061	-0.001	0.063
mutagenesis_type[T.T373A]	-0.0114	0.016	-0.700	0.485	-0.044	0.021
mutagenesis_type[T.T373A_S392A]	-0.0105	0.016	-0.649	0.517	-0.043	0.022
mutagenesis_type[T.all_three]	0.0089	0.016	0.550	0.583	-0.023	0.041
mutagenesis_type[T.both]	0.0207	0.016	1.273	0.206	-0.011	0.053
mutagenesis_type[T.none]	-0.0016	0.016	-0.097	0.923	-0.034	0.031
mutagenesis_type[T.serines]	0.0188	0.016	1.154	0.251	-0.013	0.051
KR_tally	0.0194	0.001	21.990	0.000	0.018	0.021
Omnibus:	5.357	Durbin-Watson:	0.783			
Prob(Omnibus):	0.069	Jarque-Bera (JB):	4.300			
Skew:	-0.342	Prob(JB):	0.117			
Kurtosis:	2.417	Cond. No.	74.8			
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Figure 20c. Multivariate OLS analyses for data in Figure 20a.

OLS Regression Results						
Dep. Variable:	normalized_secreted_score	R-squared:	0.625			
Model:	OLS	Adj. R-squared:	0.600			
Method:	Least Squares	F-statistic:	24.81			
Date:	Wed	15 May 2024	Prob (F-statistic):	4.27e-22		
Time:	11:41:31	Log-Likelihood:	57.193			
No. Observations:	128	AIC:	-96.39			
Df Residuals:	119	BIC:	-70.72			
Df Model:	8					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.5004	0.046	10.809	0.000	0.409	0.592
mutagenesis_type[T.S399A]	-0.0997	0.057	-1.757	0.082	-0.212	0.013
mutagenesis_type[T.S399A_T373A]	-0.1668	0.057	-2.939	0.004	-0.279	-0.054
mutagenesis_type[T.T373A]	-0.1723	0.057	-3.036	0.003	-0.285	-0.060
mutagenesis_type[T.T373A_S392A]	-0.1273	0.057	-2.242	0.027	-0.240	-0.015
mutagenesis_type[T.all_three]	-0.1669	0.057	-2.940	0.004	-0.279	-0.054
mutagenesis_type[T.none]	-0.0873	0.057	-1.538	0.127	-0.200	0.025
mutagenesis_type[T.serines]	-0.0471	0.057	-0.831	0.408	-0.160	0.065
KR_tally	0.0415	0.003	13.474	0.000	0.035	0.048
Omnibus:	0.374	Durbin-Watson:	1.643			
Prob(Omnibus):	0.830	Jarque-Bera (JB):	0.535			
Skew:	-0.057	Prob(JB):	0.765			
Kurtosis:	2.705	Cond. No.	74.8			
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Figure 20d. Multivariate OLS analyses for data shown in Figure 20b.

CHAPTER 4

DISCUSSION

Combining the preliminary Western blot data with the predictions of both MULocDeep and DeepLoc, a trend has been established between the number of positively-charged motifs and levels of CALR del52 secretion. Although the DeepLoc trends were much more conservative than the MULocDeep ones, both displayed a decline in secretion when the motifs were lost. This connection between the presence of positively-charged residues, along with the lack of negatively-charged residues like glutamic acid or aspartic acid, is also observed in the decreased levels of secretion in CALR ins5 compared to del52. There is also a recovery in secretion when these residues are mutated to alanines (see Figure 16b). The addition of these motifs into the CALR WT sequence did increase the level of secretion, but this was still lower than del52, probably due to the countering effect of the KDEL anchor sequence. Even the presence of multiple KDEL residues was not sufficient to completely halt secretion of CALR del52, suggesting that there is a balancing act between the retention-oriented KDEL and pro-secretion positively-charged residues. This may explain the secretion of CALR WT upon immunogenic cell death and the role of the KKRK motif in this process [43]. This counterbalance is observed to some extent in ins5 where some of the negatively charged residues (E/D) in the WT sequence are still present, and there is decreased secretion of ins5 compared to del52 (Figure 16c). Similar results have also been found in literature comparing the impact of these two mutations on patient outcomes [90]. This counterbalance is weakened when the E/D residues are mutated to alanines, in the case of ins5 (Figure 16c) or WT, where KKRK gains function as a pro-extracellular motif (see Figure 13c, 13d, and Table 9b).

As seen with the bar plots showing the secretion levels for addition or removal of predicted secretion motifs, they individually do not affect the activity of mutant CALR, but in unison, they have a significant impact on its localization. This is also seen in the preliminary western blot data (see Figure 9c) where the intensity of the band, representing intracellular CALR, slowly increases to the level seen in the CALR without exon 9 (covering the C-terminus). Apart from localization of the protein, the motifs also affect PTMs of nearby residues on the C-terminus, particularly S392, S399, and T373. Phosphorylation has been associated with transport to the nucleus [120], and ELM had identified several potential nuclear localization signals (NLS), some of which are our predicted secretion motifs (see Table 4c). However, some of these motifs have also been identified as extracellular motifs, suggesting two main routes for the C-terminus to undergo. Phosphorylation of these serine/threonine residues may be required for nuclear localization, as this type of PTM is generally associated with nuclear import, but further wet-lab experimentation is required. In terms of affecting secretion, mutating these residues to alanines did not significantly impact predicted secretion levels (see Figures 20a-20b).

In addition to the presence or lack of positively-charged residues, there was also the question of the 3-D structure of CALR and how it is affected by either type of frameshift mutation. It had been hypothesized that the C-terminus for ins5 could have rearranged in such a way that the relatively negatively charged aspartic and glutamic acids interfered with the predicted secretion motifs. This was seen in the close vicinity of the (K)KRK motif, where the side chains of the nearby aspartic or glutamic acids were within interacting distance (approximately five angstroms), thus preventing the motif from being in an accessible, and therefore active, conformation. This also explains how CALR ins5 is still able to be secreted despite the presence of these negatively charged residues, albeit to a lesser extent

than del52. There may be another role for these residues that hinder secretion such as binding to another protein that ultimately have the same effect, but this requires further investigation to determine as there is no data from the ELM server unlike the predicted secretion motifs.

CHAPTER 5

FUTURE DIRECTIONS

Given the breadth of *in silico* analyses, there is still a need for wet-lab experimentation that provides a quantitative answer of how either CALR del52 or ins5 are able to function, if at all, when their secretion motifs are mutated to alanines. The impact of the predicting secretion signals should be assessed by multiple methods. First stable cell lines expressing the mutants should be generated and subjected to confocal microscopy to study the subcellular localization of the mutants using specific markers for ER, ERGIC and GOLGI compartments. Secretion will be measured using ELISA with an antibody recognizing the N- terminal domain. Additionally, cell proliferation assays will be performed to assess the proliferative potential of the mutants and compare it to their secretion.

Another future direction is an emphasis on the possible nuclear localization of mutant CALR (either del52 or ins5) and how that may influence cell proliferation. Based on the ELM results, there is likely some nuclear localization of mutant CALR thanks to the positively charged prediction motifs, and it also has been documented in previous literature [69, 107]. However, the exact role of mutant CALR inside the nucleus remains elusive and needs to be explored further in future research. Similar to measurements in secretion, the accumulated loss of these motifs will be examined to see how it affects nuclear localization and the wider ramifications may be determined from that point.

Taking all the information about the qualities of these motifs into consideration, an eventual avenue of research will be to develop small molecules that would bind to these signals and mask them to prevent secretion and shuttling to the nucleus, thus preventing the expansion of MPNs, which themselves are manageable, and possible leukemic transformation to sAML, which has a far worse outcome for the patient and as of the time of writing, has no particularly effective means of treatment.

REFERENCES

1. Jagannathan-Bogdan M, Zon LI (2013) Hematopoiesis. *Development* 140:2463–2467. <https://doi.org/10.1242/DEV.083147>
2. Kaushansky K, Lok S, Holly RD, Broudy VC, Lin N, Bailey MC, Forstrom JW, Buddle MM, Oort PJ, Hagen FS, Roth GJ, Papayannopoulou T, Foster DC (1994) Promotion of megakaryocyte progenitor expansion and differentiation by the c-Mpl ligand thrombopoietin. *Nat* 1994 3696481 369:568–571. <https://doi.org/10.1038/369568a0>
3. Weeks LD, Niroula A, Neuberg D, Wong W, Lindsley RC, Luskin MR, Berliner N, Stone RM, DeAngelo DJ, Soiffer RJ, Uddin MM, Griffin G, Vlasschaert C, Gibson CJ, Jaiswal S, Bick AG, Malcovati L, Natarajan P, Ebert BL (2023) Prediction of Risk for Myeloid Malignancy in Clonal Hematopoiesis. *NEJM Evid* 2:.. https://doi.org/10.1056/EVIDOA2200310/SUPPL_FILE/EVIDOA2200310_DATA-SHARING.PDF
4. Hasserjian RP, Steensma DP, Graubert TA, Ebert BL (2020) Clonal hematopoiesis and measurable residual disease assessment in acute myeloid leukemia. *Blood* 135:1729–1738
5. McMullin MF, Anderson LA (2020) Aetiology of Myeloproliferative Neoplasms. *Cancers (Basel)* 12:1–11. <https://doi.org/10.3390/CANCERS12071810>
6. Tefferi A, Pardanani A (2015) Myeloproliferative Neoplasms: A Contemporary Review. *JAMA Oncol* 1:97–105. <https://doi.org/10.1001/JAMAONCOL.2015.89>
7. Tremblay D, Yacoub A, Hoffman R (2021) Overview of MPNs: History, Pathogenesis, Diagnostic Criteria, and Complications. *Hematol Oncol Clin North Am* 35:159. <https://doi.org/10.1016/J.HOC.2020.12.001>

8. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M, Vardiman JW (2016) The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 127:2391–2405. <https://doi.org/10.1182/BLOOD-2016-03-643544>
9. Bose P, Verstovsek S (2016) Prognosis of primary myelofibrosis in the genomic era. *Clin Lymphoma Myeloma Leuk* 16 Suppl:S105. <https://doi.org/10.1016/J.CLML.2016.02.031>
10. Baumeister J, Chatain N, Sofias AM, Lammers T, Koschmieder S (2021) Progression of Myeloproliferative Neoplasms (MPN): Diagnostic and Therapeutic Perspectives. *Cells* 2021, Vol 10, Page 3551 10:3551. <https://doi.org/10.3390/CELLS10123551>
11. Lussana F, Rambaldi A (2017) Inflammation and myeloproliferative neoplasms. *J Autoimmun* 85:58–63. <https://doi.org/10.1016/J.JAUT.2017.06.010>
12. Fisher DAC, Fowles JS, Zhou A, Oh ST (2021) Inflammatory Pathophysiology as a Contributor to Myeloproliferative Neoplasms. *Front Immunol* 12:683401. <https://doi.org/10.3389/FIMMU.2021.683401/BIBTEX>
13. Geyer HL, Dueck AC, Scherber RM, Mesa RA (2015) Impact of Inflammation on Myeloproliferative Neoplasm Symptom Development. *Mediators Inflamm* 2015:. <https://doi.org/10.1155/2015/284706>
14. Frederiksen H, Szépligeti S, Bak M, Ghanima W, Hasselbalch HC, Christiansen CF (2019) Vascular diseases in patients with chronic myeloproliferative neoplasms – Impact of comorbidity. *Clin Epidemiol* 11:955–967. <https://doi.org/10.2147/CLEP.S216787>
15. Harutyunyan AS, Kralovics R (2012) Role of germ-line Genetic Factors in MPN Pathogenesis. *Hematol Oncol Clin North Am* 26:1037–1051. <https://doi.org/10.1016/J.HOC.2012.07.005>
16. McMullin MF (2019) Diagnostic workflow for hereditary erythrocytosis and thrombocytosis. *Hematol Am Soc Hematol Educ Progr* 2019:391. <https://doi.org/10.1182/HEMATOLOGY.2019000047>
17. Ramanathan G, Hoover BM, Fleischman AG (2020) Impact of Host, Lifestyle and Environmental Factors in the Pathogenesis of MPN. *Cancers* 2020, Vol 12, Page 2038 12:2038. <https://doi.org/10.3390/CANCERS12082038>
18. Mehta J, Wang H, Iqbal SU, Mesa R (2014) Epidemiology of myeloproliferative neoplasms in the United States. *Leuk Lymphoma* 55:595–600. https://doi.org/10.3109/10428194.2013.813500/SUPPL_FILE/DISCLOSURE.ZIP

19. Price GL, Davis KL, Karve S, Pohl G, Walgren RA (2014) Survival Patterns in United States (US) Medicare Enrollees with Non-CML Myeloproliferative Neoplasms (MPN). *PLoS One* 9:e90299. <https://doi.org/10.1371/JOURNAL.PONE.0090299>
20. Hong J, Lee JH, Byun JM, Lee JY, Koh Y, Shin DY, Lee JO, Hwang SM, Choi HS, Kim I, Yoon SS, Bang SM (2019) Risk of disease transformation and second primary solid tumors in patients with myeloproliferative neoplasms. *Blood Adv* 3:3700–3708. <https://doi.org/10.1182/BLOODADVANCES.2019000655>
21. Verstovsek S, Yu J, Scherber RM, Verma S, Dieyi C, Chen CC, Parasuraman S (2022) Changes in the incidence and overall survival of patients with myeloproliferative neoplasms between 2002 and 2016 in the United States. *Leuk Lymphoma* 63:694–702. <https://doi.org/10.1080/10428194.2021.1992756>
22. Paz DL, Jouanneau-Courville R, Riou J, Ianotto JC, Boyer F, Chauveau A, Renard M, Chomel JC, Cayssials E, Gallego-Hernanz MP, Pastoret C, Murati A, Courtier F, Rousselet MC, Quintin-Roue I, Cottin L, Orvain C, Thepot S, Chretien JM, Delneste Y, Ifrah N, Blanchet O, Hunault-Berger M, Lippert E, Ugo V (2020) Leukemic evolution of polycythemia vera and essential thrombocythemia: genomic profiles predict time to transformation. *Blood Adv* 4:4887–4897. <https://doi.org/10.1182/BLOODADVANCES.2020002271>
23. Rontauroli S, Carretta C, Parenti S, Bertesi M, Manfredini R (2022) Novel Molecular Insights into Leukemic Evolution of Myeloproliferative Neoplasms: A Single Cell Perspective. *Int. J. Mol. Sci.* 23
24. Ding Y, Harada Y, Imagawa J, Kimura A, Harada H (2009) AML1/RUNX1 point mutation possibly promotes leukemic transformation in myeloproliferative neoplasms. *Blood* 114:5201–5205. <https://doi.org/10.1182/BLOOD-2009-06-223982>
25. Rawlings JS, Rosler KM, Harrison DA (2004) The JAK/STAT signaling pathway. *J Cell Sci* 117:1281–1283. <https://doi.org/10.1242/JCS.00963>
26. Aaronson DS, Horvath CM (2002) A road map for those who don't know JAK-STAT. *Science* (80-) 296:1653–1655. <https://doi.org/10.1126/SCIENCE.1071545/ASSET/F0B466F3-364A-4A06-96A1-AF7833D916CE/ASSETS/GRAPHIC/SE2020533001.JPEG>
27. Kiu, H., & Nicholson SE (2012) Biology and significance of the JAK/STAT signalling pathways. *Growth Factors* 30:88–106. <https://doi.org/10.3109/08977194.2012.660936>
28. Kwok RPS, Lundblad JR, Chrivia JC, Richards JP, Bächinger HP, Brennan RG, Roberts SGE, Green MR, Goodman RH (1994) Nuclear protein CBP is a coactivator for the transcription factor CREB. *Nat* 1994 370:223–226. <https://doi.org/10.1038/370223a0>

29. Hu X, li J, Fu M, Zhao X, Wang W (2021) The JAK/STAT signaling pathway: from bench to clinic. *Signal Transduct Target Ther* 2021 6:1–33. <https://doi.org/10.1038/s41392-021-00791-1>
30. Owen KL, Brockwell NK, Parker BS (2019) JAK-STAT Signaling: A Double-Edged Sword of Immune Regulation and Cancer Progression. *Cancers* 2019, Vol 11, Page 2002 11:2002. <https://doi.org/10.3390/CANCERS11122002>
31. Han L, Schubert C, Köhler J, Schemionek M, Isfort S, Brümmendorf TH, Koschmieder S, Chatain N (2016) Calreticulin-mutant proteins induce megakaryocytic signaling to transform hematopoietic cells and undergo accelerated degradation and Golgi-mediated secretion. *J Hematol Oncol* 9:1–14. <https://doi.org/10.1186/S13045-016-0275-0/FIGURES/6>
32. Murray PJ (2007) The JAK-STAT Signaling Pathway: Input and Output Integration. *J Immunol* 178:2623–2629. <https://doi.org/10.4049/jimmunol.178.5.2623>
33. Baker, S., Rane, S., Reddy E (2007) Hematopoietic cytokine receptor signaling. *Oncogene* 26:6724–6737. <https://doi.org/10.1038/sj.onc.1210757>
34. Fasouli, E. S., Katsantoni E (2021) JAK-STAT in early hematopoiesis and leukemia. *Front Cell Dev Biol* 9:669363
35. Vainchenker W, Constantinescu SN (2012) JAK/STAT signaling in hematological malignancies. *Oncogene* 2013 3221 32:2601–2613. <https://doi.org/10.1038/onc.2012.347>
36. Alex J. Tipping, Cristina Pina, Anders Castor, Dengli Hong, Neil P. Rodrigues, Lorenza Lazzari, Gillian E. May, Sten Eirik W. Jacobsen TE (2009) High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle. *Blood* 113:2661–2672
37. Li J (2011) Quiescence regulators for hematopoietic stem cell. *Exp Hematol* 39:511–520. <https://doi.org/10.1016/J.EXPHEM.2011.01.008>
38. Skoda RC, Duek A, Grisouard J (2015) Pathogenesis of myeloproliferative neoplasms. *Exp Hematol* 43:599–608. <https://doi.org/10.1016/J.EXPHEM.2015.06.007>
39. Yuko Kato, Atsushi Iwama, Yuko Tadokoro, Kazuya Shimoda, Mayu Minoguchi, Shizuo Akira, Minoru Tanaka, Atsushi Miyajima, Toshio Kitamura HN (2005) Selective activation of STAT5 unveils its role in stem cell self-renewal in normal and leukemic hematopoiesis. *J Exp Med* 202:169–179. <https://doi.org/10.1084/jem.20042541>
40. Brendan J. Jenkins, Andrew W. Roberts, Claire J. Greenhill, Meri Najdovska, Therese Lundgren-May, Lorraine Robb, Dianne Grail ME (2007) Pathologic consequences of STAT3

hyperactivation by IL-6 and IL-11 during hematopoiesis and lymphopoiesis. *Blood* 109:2380–2388

41. Alister C. Ward, Ivo Touw AY (2000) The Jak-Stat pathway in normal and perturbed hematopoiesis. *Blood* 95:19–29. <https://doi.org/10.1182/blood.V95.1.19>
42. O’Sullivan JM, Harrison CN (2017) JAK-STAT signaling in the therapeutic landscape of myeloproliferative neoplasms. *Mol Cell Endocrinol* 451:71–79. <https://doi.org/10.1016/J.MCE.2017.01.050>
43. Basso-Valentina F, Levy G, Varghese LN, Oufadem M, Neven B, Boussard C, Balayn N, Marty C, Vainchenker W, Plo I, Ballerini P, Constantinescu SN, Favier R, Raslova H (2021) CALR mutant protein rescues the response of MPL p.R464G variant associated with CAMT to eltrombopag. *Blood* 138:480–485. <https://doi.org/10.1182/BLOOD.2020010567>
44. Constantinescu SN, Vainchenker W, Levy G, Papadopoulos N (2021) Functional Consequences of Mutations in Myeloproliferative Neoplasms. *HemaSphere* 5:E578. <https://doi.org/10.1097/HS9.0000000000000578>
45. Orlova A, Wagner C, De Araujo ED, Bajusz D, Neubauer HA, Herling M, Gunning PT, Keserü GM, Moriggl R (2019) Direct Targeting Options for STAT3 and STAT5 in Cancer. *Cancers* 2019, Vol 11, Page 1930 11:1930. <https://doi.org/10.3390/CANCERS11121930>
46. Verstovsek S, Mesa RA, Gotlib J, Levy RS, Gupta V, Dipersio JF, Catalano J V., Deininger M, Miller C, Silver RT, Talpaz M, Winton EF, Harvey JH, Arcasoy MO, Hexner E, Lyons RM, Paquette R, Raza A, Vaddi K, Erickson-Viitanen S, Sun W, Sandor V, Kantarjian HM (2013) The clinical benefit of ruxolitinib across patient subgroups: analysis of a placebo-controlled, Phase III study in patients with myelofibrosis. *Br J Haematol* 161:508–516. <https://doi.org/10.1111/BJH.12274>
47. Dasgupta M, Dermawan JKT, Willard B, Stark GR (2015) STAT3-driven transcription depends upon the dimethylation of K49 by EZH2. *Proc Natl Acad Sci U S A* 112:3985–3990. https://doi.org/10.1073/PNAS.1503152112/SUPPL_FILE/PNAS.201503152SI.PDF
48. Balliu M, Calabresi L, Bartalucci N, Romagnoli S, Maggi L, Manfredini R, Lulli M, Guglielmelli P, Vannucchi AM (2021) Activated IL-6 signaling contributes to the pathogenesis of, and is a novel therapeutic target for, CALR-mutated MPNs. *Blood Adv* 5:2184–2195. <https://doi.org/10.1182/BLOODADVANCES.2020003291>
49. Bandaranayake RM, Ungureanu D, Shan Y, Shaw DE, Silvennoinen O, Hubbard SR (2012) Crystal structures of the JAK2 pseudokinase domain and the pathogenic mutant V617F. *Nat Struct Mol Biol* 2012 198 19:754–759. <https://doi.org/10.1038/nsmb.2348>

50. Benton CB, Boddu PC, DiNardo CD, Bose P, Wang F, Assi R, Pemmaraju N, Devendra KC, Pierce S, Patel K, Konopleva M, Ravandi F, Garcia-Manero G, Kadia TM, Cortes J, Kantarjian HM, Andreeff M, Verstovsek S (2019) Janus kinase 2 variants associated with the transformation of myeloproliferative neoplasms into acute myeloid leukemia. *Cancer* 125:1855–1866. <https://doi.org/10.1002/CNCR.31986>
51. Toms A V., Deshpande A, McNally R, Jeong Y, Rogers JM, Kim CU, Gruner SM, Ficarro SB, Marto JA, Sattler M, Griffin JD, Eck MJ (2013) Structure of a pseudokinase-domain switch that controls oncogenic activation of Jak kinases. *Nat Struct Mol Biol* 20:1221–1223. <https://doi.org/10.1038/nsmb.2673>
52. Shan Y, Gnanasambandan K, Ungureanu D, Kim ET, Hammarén H, Yamashita K, Silvennoinen O, Shaw DE, Hubbard SR (2014) Molecular basis for pseudokinase-dependent autoinhibition of JAK2 tyrosine kinase. *Nat Struct Mol Biol* 21:579–584. <https://doi.org/10.1038/nsmb.2849>
53. Tefferi A, Guglielmelli P, Larson DR, Finke C, Wassie EA, Pieri L, Gangat N, Fjerza R, Belachew AA, Lasho TL, Ketterling RP, Hanson CA, Rambaldi A, Finazzi G, Thiele J, Barbui T, Pardanani A, Vannucchi AM (2014) Long-term survival and blast transformation in molecularly annotated essential thrombocythemia, polycythemia vera, and myelofibrosis. *Blood* 124:2507–2513. <https://doi.org/10.1182/BLOOD-2014-05-579136>
54. Vainchenker W, Kralovics R (2017) Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms. *Blood* 129:667–679. <https://doi.org/10.1182/BLOOD-2016-10-695940>
55. Hermouet S, Bigot-Corbel E, Gardie B (2015) Pathogenesis of Myeloproliferative Neoplasms: Role and Mechanisms of Chronic Inflammation. *Mediators Inflamm* 2015:. <https://doi.org/10.1155/2015/145293>
56. Chen E, Mullally A (2014) How does JAK2V617F contribute to the pathogenesis of myeloproliferative neoplasms? *Hematology* 2014:268–276. <https://doi.org/10.1182/ASHEDUCATION-2014.1.268>
57. Grabek J, Straube J, Bywater M, Lane SW (2020) MPN: The Molecular Drivers of Disease Initiation, Progression and Transformation and their Effect on Treatment. *Cells* 2020, Vol 9, Page 1901 9:1901. <https://doi.org/10.3390/CELLS9081901>
58. Stegelmann F, Bullinger L, Schlenk RF, Paschka P, Griesshammer M, Blersch C, Kuhn S, Schauer S, Döhner H, Döhner K (2011) DNMT3A mutations in myeloproliferative neoplasms. *Leuk* 25:1217–1219. <https://doi.org/10.1038/leu.2011.77>

59. Pardanani AD, Levine RL, Lasho T, Pikman Y, Mesa RA, Wadleigh M, Steensma DP, Elliott MA, Wolanskyj AP, Hogan WJ, McClure RF, Litzow MR, Gilliland DG, Tefferi A (2006) MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. *Blood* 108:3472–3476. <https://doi.org/10.1182/BLOOD-2006-04-018879>
60. Xie J, Chen X, Gao F, Hou R, Tian T, Zhang Y, Fan L, Hu J, Zhu G, Yang W, Wang H (2019) Two activating mutations of MPL in triple-negative myeloproliferative neoplasms. *Cancer Med* 8:5254–5263. <https://doi.org/10.1002/CAM4.2387>
61. Varghese LN, Defour JP, Pecquet C, Constantinescu SN (2017) The thrombopoietin receptor: Structural basis of traffic and activation by ligand, mutations, agonists, and mutated calreticulin. *Front Endocrinol (Lausanne)* 8:250740. <https://doi.org/10.3389/FENDO.2017.00059/BIBTEX>
62. Varghese LN, Carreño-Tarragona G, Levy G, Gutiérrez-López de Ocariz X, Rapado I, Martínez-López J, Ayala R, Constantinescu SN (2021) MPL S505C enhances driver mutations at W515 in essential thrombocythemia. *Blood Cancer J* 2021 1111 11:1–3. <https://doi.org/10.1038/s41408-021-00583-4>
63. Levy G, Carillo S, Papoular B, Cassinat B, Zini JM, Leroy E, Varghese LN, Chachoua I, Defour JP, Smith SO, Constantinescu SN (2020) MPL mutations in essential thrombocythemia uncover a common path of activation with eltrombopag dependent on W491. *Blood* 135:948–953. <https://doi.org/10.1182/BLOOD.2019003240>
64. Bridgford JL, Lee SM, Lee CMM, Guglielmelli P, Rumi E, Pietra D, Wilcox S, Chhabra Y, Rubin AF, Cazzola M, Vannucchi AM, Brooks AJ, Call ME, Call MJ (2020) Novel drivers and modifiers of MPL-dependent oncogenic transformation identified by deep mutational scanning. *Blood* 135:287–292. <https://doi.org/10.1182/BLOOD.2019002561>
65. Guglielmelli P, Calabresi L (2021) The MPL mutation. *Int Rev Cell Mol Biol* 365:163–178. <https://doi.org/10.1016/BS.IRCMB.2021.09.003>
66. Chachoua I, Pecquet C, El-Khoury M, Nivarthi H, Albu RI, Marty C, Gryshkova V, Defour JP, Vertenoeil G, Ngo A, Koay A, Raslova H, Courtoy PJ, Choong ML, Plo I, Vainchenker W, Kralovics R, Constantinescu SN (2016) Thrombopoietin receptor activation by myeloproliferative neoplasm associated calreticulin mutants. *Blood* 127:1325–1335. <https://doi.org/10.1182/BLOOD-2015-11-681932>
67. Beer PA, Ortmann CA, Stegelmann F, Guglielmelli P, Reilly JT, Larsen TS, Hasselbalch HC, Vannucchi AM, Möller P, Döhner K, Green AR (2010) Molecular mechanisms associated with leukemic transformation of MPL-mutant myeloproliferative neoplasms. *Haematologica* 95:2153. <https://doi.org/10.3324/HAEMATOL.2010.029306>

68. Marty C, Pecquet C, Nivarthi H, El-Khoury M, Chachoua I, Tulliez M, Villeval JL, Raslova H, Kralovics R, Constantinescu SN, Plo I, Vainchenker W (2016) Calreticulin mutants in mice induce an MPL-dependent thrombocytosis with frequent progression to myelofibrosis. *Blood* 127:1317–1324. <https://doi.org/10.1182/BLOOD-2015-11-679571>
69. Masubuchi N, Araki M, Yang Y, Hayashi E, Imai M, Edahiro Y, Hironaka Y, Mizukami Y, Kihara Y, Takei H, Nudejima M, Koike M, Ohsaka A, Komatsu N (2019) Mutant calreticulin interacts with MPL in the secretion pathway for activation on the cell surface. *Leuk* 2019 342 34:499–509. <https://doi.org/10.1038/s41375-019-0564-z>
70. Araki M, Yang Y, Imai M, Mizukami Y, Kihara Y, Sunami Y, Masubuchi N, Edahiro Y, Hironaka Y, Osaga S, Ohsaka A, Komatsu N (2018) Homomultimerization of mutant calreticulin is a prerequisite for MPL binding and activation. *Leuk* 2018 331 33:122–131. <https://doi.org/10.1038/s41375-018-0181-2>
71. Krause KH, Michalak M (1997) Calreticulin. *Cell* 88:439–443. [https://doi.org/10.1016/S0092-8674\(00\)81884-X](https://doi.org/10.1016/S0092-8674(00)81884-X)
72. Coppelino MG, Dedhar S (1998) Calreticulin. *Int J Biochem Cell Biol* 30:553–558. [https://doi.org/10.1016/S1357-2725\(97\)00153-2](https://doi.org/10.1016/S1357-2725(97)00153-2)
73. Lu YC, Weng WC, Lee H (2015) Functional roles of calreticulin in cancer biology. *Biomed Res Int* 2015:.. <https://doi.org/10.1155/2015/526524>
74. Bouvier M, Stafford WF (2000) Probing the three-dimensional structure of human calreticulin. *Biochemistry* 39:14950–14959. <https://doi.org/10.1021/BI0019545/ASSET/IMAGES/LARGE/BI0019545F000006.JPEG>
75. Michalak M, Parker JMR, Opas M (2002) Ca²⁺ signaling and calcium binding chaperones of the endoplasmic reticulum. *Cell Calcium* 32:269–278. <https://doi.org/10.1016/S0143416002001884>
76. Marek Michalak, Jody Groenendyk, Eva Szabo, Leslie I. Gold MO (2009) Calreticulin, a multi-process calcium-buffering chaperone of the endoplasmic reticulum. *Biochem J* 417:651–666. <https://doi.org/10.1042/BJ20081847>
77. Coppelino MG, Woodside MJ, Demarex N, Grinstein S, St-Arnaud R, Dedhar S (1997) Calreticulin is essential for integrin-mediated calcium signalling and cell adhesion. *Nat* 1997 386:843–847. <https://doi.org/10.1038/386843a0>

78. Maria Villagomez, Eva Szabo, Alexey Podcheko, Tianshu Feng, Sylvia Papp MO (2009) Calreticulin and focal-contact-dependent adhesion. *Biochem Cell Biol* 87:545–556. <https://doi.org/10.1139/O09-016>
79. Chakrabarti A, Chen AW, Varner JD (2011) A review of the mammalian unfolded protein response. *Biotechnol Bioeng* 108:2777–2793. <https://doi.org/10.1002/BIT.23282>
80. Yang, Y., Ma, F., Liu, Z., Su, Q., Liu, Y., Liu, Z., & Li Y (2019) The ER-localized Ca²⁺-binding protein calreticulin couples ER stress to autophagy by associating with microtubule-associated protein 1A/1B light chain 3. *J Biol Chem* 294:772–782. <https://doi.org/10.1074/jbc.RA118.005166>
81. Theocharis Panaretakis, Oliver Kepp, Ulf Brockmeier, Antoine Tesniere, Ann Charlotte Bjorklund, Daniel C Chapman, Michael Durchschlag, Nicholas Joza, Gérard Pierron, Peter van Endert, Junying Yuan, Laurence Zitvogel, Frank Madeo, David B Williams and GK (2009) Mechanisms of pre-apoptotic calreticulin exposure in immunogenic cell death. *EMBO J* 28:578–590. <https://doi.org/10.1038/emboj.2009.1>
82. Martins I, Kepp O, Galluzzi L, Senovilla L, Schlemmer F, Adjemian S, Menger L, Michaud M, Zitvogel L, Kroemer G (2010) Surface-exposed calreticulin in the interaction between dying cells and phagocytes. *Ann N Y Acad Sci* 1209:77–82. <https://doi.org/10.1111/J.1749-6632.2010.05740.X>
83. Kepp O, Liu P, Zhao L, Plo I, Kroemer G (2020) Surface-exposed and soluble calreticulin: conflicting biomarkers for cancer prognosis. <https://doi.org/10.1080/2162402X.2020.1792037>
84. Gold LI, Rahman M, Blechman KM, Greives MR, Churgin S, Michaels J, Callaghan MJ, Cardwell NL, Pollins AC, Michalak M, Siebert JW, Levine JP, Gurtner GC, Nanney LB, Galiano RD, Cadacio CL (2006) Overview of the Role for Calreticulin in the Enhancement of Wound Healing through Multiple Biological Effects. *J Investig Dermatology Symp Proc* 11:57–65. <https://doi.org/10.1038/SJ.JIDSYMP.5650011>
85. Marneth AE, Mullally A (2020) The Molecular Genetics of Myeloproliferative Neoplasms. *Cold Spring Harb Perspect Med* 10:. <https://doi.org/10.1101/CSHPERSPECT.A034876>
86. Nangalia J, Massie CE, Baxter EJ, Nice FL, Gundem G, Wedge DC, Avezov E, Li J, Kollmann K, Kent DG, Aziz A, Godfrey AL, Hinton J, Martincorena I, Van Loo P, Jones AV, Guglielmelli P, Tarpey P, Harding HP, Fitzpatrick JD, Goudie CT, Ortmann CA, Loughran SJ, Raine K, Jones DR, Butler AP, Teague JW, O'Meara S, McLaren S, Bianchi M, Silber Y, Dimitropoulou D, Bloxham D, Mudie L, Maddison M, Robinson B, Keohane C, Maclean C, Hill K, Orchard K, Tauro S, Du M-Q, Greaves M, Bowen D, Huntly BJP, Harrison CN, Cross NCP, Ron D,

- Vannucchi AM, Papaemmanuil E, Campbell PJ, Green AR (2013) Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med* 369:2391–2405. <https://doi.org/10.1056/NEJMOA1312542>
87. Imai M, Araki M, Komatsu N (2017) Somatic mutations of calreticulin in myeloproliferative neoplasms. *Int J Hematol* 105:743–747. <https://doi.org/10.1007/S12185-017-2246-9/FIGURES/2>
 88. Schürch PM, Malinowska L, Hleihil M, Losa M, Hofstetter MC, Wildschut MH, Lysenko V, Lakkaraju AK, Maat CA, Benke D, Aguzzi A, Wollscheid B, Picotti P, Theodorides AP (2022) Calreticulin mutations affect its chaperone function and perturb the glycoproteome Graphical abstract. *Cell Rep* 41:111689. <https://doi.org/10.1016/j.celrep.2022.111689>
 89. Klampfl T, Gisslinger H, Harutyunyan AS, Nivarthi H, Rumi E, Milosevic JD, Them NCC, Berg T, Gisslinger B, Pietra D, Chen D, Vladimer GI, Bagienski K, Milanesi C, Casetti IC, Sant'Antonio E, Ferretti V, Elena C, Schischlik F, Cleary C, Six M, Schalling M, Schönegger A, Bock C, Malcovati L, Pascutto C, Superti-Furga G, Cazzola M, Kralovics R (2013) Somatic Mutations of Calreticulin in Myeloproliferative Neoplasms. *N Engl J Med* 369:2379–2390. https://doi.org/10.1056/NEJMOA1311347/SUPPL_FILE/NEJMOA1311347_DISCLOSURES.PDF
 90. Pecquet C, Papadopoulos N, Balligand T, Chachoua I, Tisserand A, Vertenoeil G, Nédélec A, Vertommen D, Roy A, Marty C, Nivarthi H, Defour J-P, El-Khoury M, Hug E, Majoros A, Xu E, Zagrijtschuk O, Fertig TE, Marta DS, Gisslinger H, Gisslinger B, Schalling M, Casetti I, Rumi E, Pietra D, Cavalloni C, Arcaini L, Cazzola M, Komatsu N, Kihara Y, Sunami Y, Edahiro Y, Araki M, Lesyk R, Buxhofer-Ausch V, Heibl S, Pasquier F, Havelange V, Plo I, Vainchenker W, Kralovics R, Constantinescu SN (2023) Secreted mutant calreticulins as rogue cytokines in myeloproliferative neoplasms. *Blood* 141:917–929. <https://doi.org/10.1182/BLOOD.2022016846>
 91. Raghavan M, Wijeyesakere SJ, Peters LR, Del Cid N (2013) Calreticulin in the immune system: Ins and outs. *Trends Immunol* 34:13–21. <https://doi.org/10.1016/J.IT.2012.08.002>
 92. Liu P, Zhao L, Kroemer G, Kepp O (2020) Secreted calreticulin mutants subvert anticancer immunosurveillance. *Oncoimmunology* 9:. <https://doi.org/10.1080/2162402X.2019.1708126>
 93. Rumi E, Pietra D, Pascutto C, Guglielmelli P, Martínez-Trillos A, Casetti I, Colomer D, Pieri L, Pratcorona M, Rotunno G, Sant'Antonio E, Bellini M, Cavalloni C, Mannarelli C, Milanesi C, Boveri E, Ferretti V, Astori C, Rosti V, Cervantes F, Barosi G, Vannucchi AM, Cazzola M,

- (2014) Clinical effect of driver mutations of JAK2, CALR, or MPL in primary myelofibrosis. *Blood* 124:1062–1069. <https://doi.org/10.1182/BLOOD-2014-05-578435>
94. Luo W, Yu Z (2015) Calreticulin (CALR) mutation in myeloproliferative neoplasms (MPNs). *Stem Cell Investig* 2:16. <https://doi.org/10.3978/J.ISSN.2306-9759.2015.08.01>
 95. Shivarov V, Ivanova M, Tiu R V. (2014) Mutated calreticulin retains structurally disordered C terminus that cannot bind Ca²⁺: some mechanistic and therapeutic implications. *Blood Cancer J* 2014 42 4:e185–e185. <https://doi.org/10.1038/bcj.2014.7>
 96. Di Buduo CA, Abbonante V, Marty C, Moccia F, Rumi E, Pietra D, Soprano PM, Lim D, Cattaneo D, Iurlo A, Gianelli U, Barosi G, Rosti V, Plo I, Cazzola M, Balduini A (2020) Defective interaction of mutant calreticulin and SOCE in megakaryocytes from patients with myeloproliferative neoplasms. *Blood* 135:133–144. <https://doi.org/10.1182/BLOOD.2019001103>
 97. Chachoua I, Pecquet C, El-Khoury M, Nivarthi H, Albu RI, Marty C, Gryshkova V, Defour JP, Vertenoeil G, Ngo A, Koay A, Raslova H, Courtoy PJ, Choong ML, Plo I, Vainchenker W, Kralovics R, Constantinescu SN (2016) Thrombopoietin receptor activation by myeloproliferative neoplasm associated calreticulin mutants. *Blood* 127:1325–1335. <https://doi.org/10.1182/BLOOD-2015-11-681932>
 98. Balligand T, Achouri Y, Pecquet C, Chachoua I, Nivarthi H, Marty C, Vainchenker W, Plo I, Kralovics R, Constantinescu SN (2016) Pathologic activation of thrombopoietin receptor and JAK2-STAT5 pathway by frameshift mutants of mouse calreticulin. *Leuk* 2016 308 30:1775–1778. <https://doi.org/10.1038/leu.2016.47>
 99. Pronier E, Cifani P, Merlinsky TR, Berman KB, Somasundara AVH, Rampal RK, LaCava J, Wei KE, Pastore F, Maag JL, Park J, Koche R, Kentsis A, Levine RL (2018) Targeting the CALR interactome in myeloproliferative neoplasms. *JCI Insight* 3:. <https://doi.org/10.1172/JCI.INSIGHT.122703>
 100. Ibarra J, Elbanna YA, Kurylowicz K, Ciboddo M, Greenbaum HS, Arellano NS, Rodriguez D, Evers M, Bock-Hughes A, Liu C, Smith Q, Lutze J, Baumeister J, Kalmer M, Olschok K, Nicholson B, Silva D, Maxwell L, Dowgielewicz J, Rumi E, Pietra D, Casetti IC, Catricala S, Koschmieder S, Gurbuxani S, Schneider RK, Oakes SA, Elf SE (2022) Type I but Not Type II Calreticulin Mutations Activate the IRE1a/XBP1 Pathway of the Unfolded Protein Response to Drive Myeloproliferative Neoplasms. *Blood Cancer Discov* 3:298–315. <https://doi.org/10.1158/2643-3230.BCD-21-0144/694332/AM/TYPE-1-BUT-NOT-TYPE-2-CALRETICULIN-MUTATIONS>

101. Simona Salati, Elena Genovese, Chiara Carretta, Roberta Zini, Niccolò Bartalucci, Zelia Prudente, Valentina Pennucci, Samantha Ruberti, Chiara Rossi, Sebastiano Rontauoli, Elena Enzo, Laura Calabresi, Manjola Balliu, Carmela Mannarelli, Elisa Bianchi, Pa RM (2019) Calreticulin Ins5 and Del52 mutations impair unfolded protein and oxidative stress responses in K562 cells expressing CALR mutants. *Sci Rep* 9:. <https://doi.org/10.1038/s41598-019-46843-z>
102. Genovese, E.; Mirabile, M.; Rontauoli, S.; Sartini, S.; Fantini, S.; Tavernari, L.; Maccaferri, M.; Guglielmelli, P.; Bianchi, E.; Parenti S. et al. (2022) The Response to Oxidative Damage Correlates with Driver Mutations and Clinical Outcome in Patients with Myelofibrosis. *Antioxidants* 11:113. <https://doi.org/10.3390/antiox11010113>
103. Banfield DK (2011) Mechanisms of Protein Retention in the Golgi. *Cold Spring Harb Perspect Biol* 3:1–14. <https://doi.org/10.1101/CSHPERSPECT.A005264>
104. Uemura S, Yoshida S, Shishido F, Inokuchi JI (2009) The Cytoplasmic Tail of GM3 Synthase Defines Its Subcellular Localization, Stability, and In Vivo Activity. *Mol Biol Cell* 20:3088. <https://doi.org/10.1091/MBC.E08-12-1219>
105. Pecquet C, Chachoua I, Roy A, Balligand T, Vertenoeil G, Leroy E, Albu RI, Defour JP, Nivarthi H, Hug E, Xu E, Ould-Amer Y, Mouton C, Colau D, Vertommen D, Shwe MM, Marty C, Plo I, Vainchenker W, Kralovics R, Constantinescu SN (2019) Calreticulin mutants as oncogenic rogue chaperones for TpoR and traffic-defective pathogenic TpoR mutants. *Blood* 133:2669–2681. <https://doi.org/10.1182/BLOOD-2018-09-874578>
106. Lasho TL, Finke CM, Tischer A, Pardanani A, Tefferi A (2018) Mayo CALR mutation type classification guide using alpha helix propensity. *Am J Hematol* 93:E128–E129. <https://doi.org/10.1002/AJH.25065>
107. Garbati MR, Welgan CA, Landefeld SH, Newell LF, Agarwal A, Dunlap JB, Chourasia TK, Lee H, Elferich J, Traer E, Rattray R, Cascio MJ, Press RD, Bagby GC, Tyner JW, Druker BJ, Dao KHT (2016) Mutant calreticulin-expressing cells induce monocyte hyperreactivity through a paracrine mechanism. *Am J Hematol* 91:211–219. <https://doi.org/10.1002/AJH.24245>
108. Kumar M, Gouw M, Michael S, Sámano-Sánchez H, Pancsa R, Glavina J, Diakogianni A, Valverde JA, Bukirova D, Signalyševa J, Palopoli N, Davey NE, Chemes LB, Gibson TJ (2020) ELM—the eukaryotic linear motif resource in 2020. *Nucleic Acids Res* 48:D296–D306. <https://doi.org/10.1093/NAR/GKZ1030>
109. Pietra D, Rumi E, Ferretti V V., Di Buduo CA, Milanesi C, Cavalloni C, Sant’Antonio E, Abbonante V, Moccia F, Casetti IC, Bellini M, Renna MC, Roncoroni E, Fugazza E, Astori C, Boveri E, Rosti V, Barosi G, Balduini A, Cazzola M (2015) Differential clinical effects of

different mutation subtypes in CALR-mutant myeloproliferative neoplasms. *Leukemia* 30:431–438. <https://doi.org/10.1038/LEU.2015.277>

110. Thummuluri V, Almagro Armenteros JJ, Johansen AR, Nielsen H, Winther O (2022) DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res* 50:W228–W234. <https://doi.org/10.1093/NAR/GKAC278>
111. Jiang Y, Wang D, Yao Y, Eubel H, Künzler P, Möller IM, Xu D (2021) MULocDeep: A deep-learning framework for protein subcellular and suborganellar localization prediction with residue-level interpretation. *Comput Struct Biotechnol J* 19:4825–4839. <https://doi.org/10.1016/J.CSBJ.2021.08.027>
112. Jiang Y, Jiang L, Akhil CS, Wang D, Zhang Z, Zhang W, Xu D (2023) MULocDeep web service for protein localization prediction and visualization at subcellular and suborganellar levels. *Nucleic Acids Res* 51:W343–W349. <https://doi.org/10.1093/NAR/GKAD374>
113. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M (2022) ColabFold: making protein folding accessible to all. *Nat Methods* 2022 19:679–682. <https://doi.org/10.1038/s41592-022-01488-1>
114. Badgett MJ, Boyes B, Orlando R (2017) Predicting the Retention Behavior of Specific O-Linked Glycopeptides. *J Biomol Tech* 28:122. <https://doi.org/10.7171/JBT.17-2803-003>
115. Wang D, Liu D, Yuchi J, He F, Jiang Y, Cai S, Li J, Xu D (2020) MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. *Nucleic Acids Res* 48:W140–W146. <https://doi.org/10.1093/NAR/GKAA275>
116. Luo F, Wang M, Liu Y, Zhao XM, Li A (2019) DeepPhos: prediction of protein phosphorylation sites with deep learning. *Bioinformatics* 35:2766–2773. <https://doi.org/10.1093/BIOINFORMATICS/BTY1051>
117. Yao K, Wu Y, Chen Q, Zhang Z, Chen X, Zhang Y (2016) The Arginine/Lysine-Rich Element within the DNA-Binding Domain Is Essential for Nuclear Localization and Function of the Intracellular Pathogen Resistance 1. *PLoS One* 11:e0162832. <https://doi.org/10.1371/JOURNAL.PONE.0162832>
118. Afshar N, Black BE, Paschal BM (2005) Retrotranslocation of the Chaperone Calreticulin from the Endoplasmic Reticulum Lumen to the Cytosol. *Mol Cell Biol* 25:8844. <https://doi.org/10.1128/MCB.25.20.8844-8853.2005>
119. Duarte JM, Sathyapriya R, Stehr H, Filippis I, Lappe M (2010) Optimal contact definition for reconstruction of Contact Maps. *BMC Bioinformatics* 11:283. <https://doi.org/10.1186/1471->

2105-11-283

120. Jans DA, Hübner S (1996) Regulation of protein transport to the nucleus: central role of phosphorylation. <https://doi.org/10.1152/physrev1996763651> 76:651–685.
<https://doi.org/10.1152/PHYSREV.1996.76.3.651>
121. Li L, Vorobyov I, Allen TW (2013) The different interactions of lysine and arginine side chains with lipid membranes. *J Phys Chem B* 117:11906–11920.
https://doi.org/10.1021/JP405418Y/SUPPL_FILE/JP405418Y_SI_001.PDF

