The *CEL* VNTR in pancreatic disease

Effects of cysteine residues in pathogenic carboxyl ester lipase variants

Miguel Ángel Juárez Garzón

This thesis is submitted in partial fulfilment of the requirements for the degree of Master in Biomedical Sciences

Department of Clinical Medicine and Department of Biomedicine.

University of Bergen

Spring 2023

Acknowledgements

The work presented in this thesis has mainly been performed at the Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen from February 2022 to May 2023. I have been affiliated with the master's program in Biomedical Sciences at the Department of Biomedicine, University of Bergen.

First, I would like to thank my supervisor, prof. Anders Molven, for letting me do this thesis under your supervision and for everything you have taught me during the thesis work. I am truly grateful for all your guidance, all the support, your kindness and for sharing your brilliant ideas with me. Thank you also for your great feedback and help during the writing process. I admire you and you have made me a better researcher.

I would also like to thank my co-supervisors, Khadija El Jellas and Janniche Torsvik. Khadija thanks for such a great project idea and your help despite the distance. Janniche thanks for your skilled guidance during the experimental part and for making sure I had everything needed to perform my best. Also, thanks to both for your nice feedback during the writing process.

Moreover, I am thankful to Solrun Steine, Louise Grevle and Hege A. Dale. Solrun, thanks for your help with DNA electrophoresis and genotyping, ordering reagents, and for your kindness. Louise, thank you for teaching me how to work with bacteria and plasmid prep kits. Hege, thank you for your technical support at the confocal microscope. Special thanks to Karianne Fjeld for helping me to find antibodies and reagents and to Rune H. Navarsete for the nice talks in the lab. To the rest of the CEL group, I am grateful for the valuable tips and amazing comments during our meetings. Furthermore, I would like to thank everyone in the Mohn Center for Diabetes Precision Medicine for introducing me to so many interesting fields.

Finally, I would like to thank my family and friends for your cheering and motivation. A special thank you goes to my parents, none of this would be possible without you. Thanks for being my biggest motivation and support and for always pushing me toward greatness.

Bergen, June 2023
Miguel Ángel Juárez Garzón.

Table of contents

Abbreviations

ATF6α	Activating transcription factor 6α
BiP	Binding immunoglobulin protein
bp	Base-pairs
CEL/CEL	Carboxyl ester lipase gene/protein
CEL-3R	Carboxyl ester lipase variant with only 3 repeats
CEL-HYB1	Carboxyl ester lipase hybrid allele 1
CEL-INS	Insertion variants of carboxyl ester lipase
CEL-MODY	Carboxyl ester lipase leading to MODY type 8
CELP	Carboxyl ester lipase pseudogene
CEL-WT	Carboxyl ester lipase wildtype
DM	Diabetes mellitus
ER	Endoplasmic reticulum
GALNT2	Polypeptide N-acetylgalactosaminyltransferase 2
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
HEK293 cells	Human embryonic kidney 293 cells
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered protein region
IRE1α	Inositol-requiring protein 1α
MODY	Maturity-onset diabetes of the young
PC	Pancreatic cancer
PDAC	Pancreatic ductal adenocarcinoma
PDI	Protein disulphide isomerase
PERK	PRKR-like ER kinase
PTM	Post-translational modifications
SC	Simple Cells
T1D	Type 1 diabetes
T2D	Type 2 diabetes
UPR	Unfolded protein response
VNTR	Variable number of tandem repeats

Abstract

Carboxyl ester lipase (CEL) is a digestive enzyme mainly expressed in the acinar cells of the pancreas. CEL hydrolyses dietary fat, cholesteryl esters and fat-soluble vitamins. The protein is structurally divided into two regions: an N-terminal globular domain and a C-terminal tail encoded by a variable number of tandem repeats (VNTR). The *CEL* VNTR is highly polymorphic and pathogenic variants of CEL (CEL-MODY and CEL-HYB1) are associated with mutations in this region. CEL-MODY is caused by a single-base pair deletion leading to a frameshift and a truncated CEL protein that introduces ten *de novo* cysteines not present in the normal protein tail. CEL-MODY causes Maturity-Onset Diabetes of the Young, type 8 (MODY8). The disease is also characterized by exocrine dysfunction and progressive pancreatic deterioration. The *CEL-HYB1* allele is a genetic risk factor for chronic pancreatitis that derives from recombination between *CEL* and its pseudogene *CELP*, leading to a short aberrant tail that contains two cysteines not found in the normal protein.

In this thesis, we aimed to gain new knowledge about the function of the CEL VNTR-encoded tail region, specifically the effect of *de novo* cysteines in the pathogenic variants. To this end, synthetic constructs of CEL-MODY and CEL-HYB1 proteins were created where cysteines were mutated into alanines and compared with the regular constructs as well as the normal CEL protein (CEL-WT).

When studying CEL-MODY, we found that mutating the cysteines to alanines generally normalized the cellular properties of this variant. Thus, the cysteines were responsible for the low secretion levels of CEL-MODY and affected cellular distribution, leading to a higher abundance in the detergent-insoluble pellet fraction. In addition, the cysteines caused changes in intracellular localization of CEL, by increasing the presence of the defective protein in the endoplasmic reticulum. Finally, the cysteine residues and their subsequent participation in disulphide bridges interfered with O-glycosylation of CEL-MODY.

When analysing CEL-HYB1 we were able to confirm previous knowledge about this pathogenic variant, such as low secretion and aggregation in the detergent-insoluble pellet fraction. However, mutating the cysteines into alanines did not change the cellular properties examined of CEL-HYB1.

In summary, we revealed clear differences between the impact of cysteines in the mucinous domain of the pathogenic CEL-MODY and CEL-HYB1 variants. We conclude that CEL-MODY pathogenicity may be driven mainly by cysteines present in the aberrant C-terminal tail, which lead to impaired secretion and changes in protein distribution, intracellular localization and O-glycosylation. In contrast, our data did not indicate an association between cysteines in the tail of CEL-HYB1 and this variant's behaviour. However, we speculate that the unusually short length of the CEL-HYB1 tail could be the main driver of pathogenicity.

1. Introduction

1.1 The Pancreas

The human pancreas is a retroperitoneal organ located in the upper abdomen, extending from the duodenum to the spleen, and it is surrounded by several important organs and blood vessels (**Figure 1A**). The pancreas is normally divided into four anatomic regions: the head, embedded within the curvature of the duodenum and closely attached by connective tissue, the neck, the body, and the tail, entering the hilum of the spleen [1]. The main pancreatic duct (duct of Wirsung) extends through the long axis of the gland to the duodenum. The duct is connected to the duodenum via the ampulla of Vater, where the pancreatic and bile ducts are joined to release their contents into the descending part of the duodenum [2, 3] (**Figure 1B**). As will be described in further detail below, the pancreas contains two different types of tissue, exocrine cells, which produce digestive enzymes in the form of pancreatic juice, and endocrine cells that secrete hormones regulating energy metabolism [3, 4] Thus, the gland can in many respects be regarded as "two organs in one".

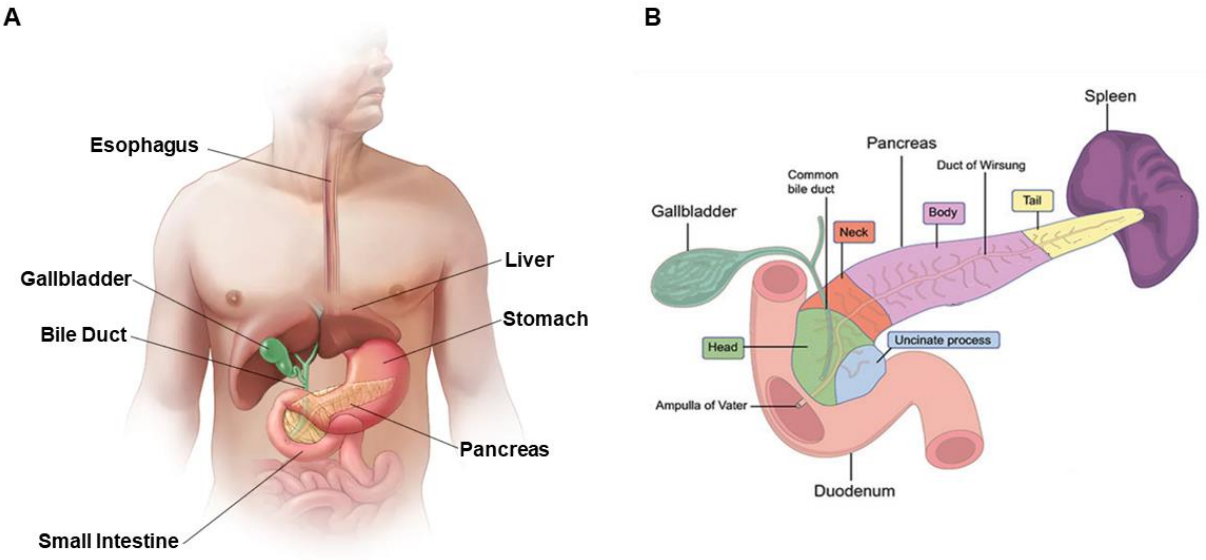

Figure 1. Anatomical overview of the human pancreas. A) The pancreas is located in the upper abdomen, below the stomach. Adapted from ref. [6] **B)** The pancreas presents four anatomical parts: the head (green), the neck (orange), the body (purple) and the tail (yellow). The pancreas releases its content into the duodenum via the ampulla of Vater. Adapted from ref. [5]

1.1.1 Endocrine Pancreas

The endocrine pancreas makes up for ~1-2 % of the total pancreatic volume [7]. It consists of the islets of Langerhans, small bead-like cell clusters scattered in the parenchyma of the pancreas [8]. The main function of the endocrine pancreas is to produce hormones that are released into the bloodstream and play a major role in the regulation of energy metabolism, in particular glucose homeostasis [3, 9]. Regarding endocrine cells, islets contain five different types of hormone-producing cells: α-cells, β-cells, δ-cells, γ-cells, and ε-cells [8] (**Figure 2**). In the human pancreas, these cells are distributed in an apparently random pattern with each type comprising a certain percentage of the total cell number (~ 60 % insulin-producing β-cells, ~30 % glucagon-producing α-cells, while the remaining ~10 % are somatostatin-producing δ-cells, pancreatic polypeptide-producing γ-cells, and ghrelin-producing ε-cells) [8]. The cellular composition and pancreatic distribution of the islets vary both between and within species [7, 10].

The main role of the endocrine pancreas is to strictly regulate glucose homeostasis. This is carried out mainly by β-cells and α-cells, through secretion of insulin and glucagon respectively [9]. Insulin stimulates the uptake and storage of glucose in peripheral tissues such as skeletal muscle, the liver and adipose tissue [9]. Glucagon, in contrast, is produced in the fasting state and it mainly serves to raise the levels of fatty acids and glucose in the blood [11]. In this way, insulin and glucagon can be regarded as counteracting hormones with respect to their effect on glucose homeostasis. The other cells of the endocrine pancreas have a complementary role. δ-cells secret somatostatin, a hormone that blocks the secretion of both insulin and glucagon from neighbouring cells [12]. γ-cells, also known as PP-cells, produce pancreatic polypeptide (PP), an inhibitor of bicarbonate, enzymes and fluid secretion from the exocrine pancreas [8]. Finally, ε-cells produce ghrelin, a modulator of insulin secretion in relation to food intake, energy expenditure and inflammation [13].

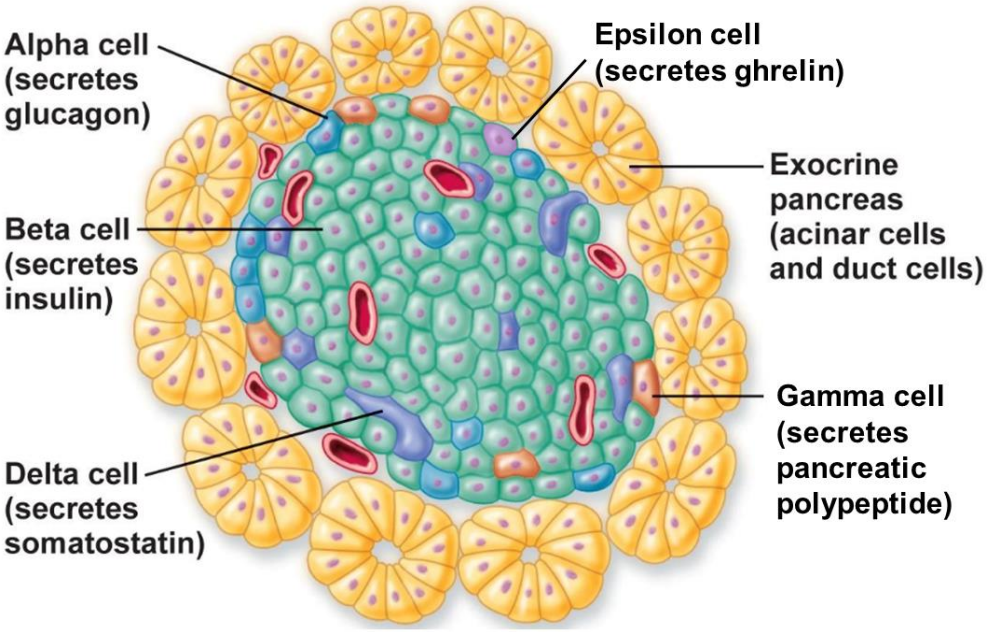

Figure 2. Endocrine cell types in the islet of Langerhans. Islets contain 5 main types of endocrine cells: beta (green), alpha (blue), delta (purple), gamma (orange) and epsilon (pink). The cells spread randomly inside the human islet and present different abundances. Islets of Langerhans are normally surrounded by acinar and duct cells (yellow). Adapted from ref. [14]

1.1.2 Exocrine pancreas

The exocrine pancreas comprises most of the total pancreatic mass (90%) and it is formed by acinar, centroacinar and ductal cells [15]. Acinar cells are the predominating cell type [16], and they are organized in lobes, which are made of smaller lobules, that in turn are made of several acini, the secretory units of the exocrine pancreas (**Figure 3**). Each acinus consists of a single layer of pyramidal acinar cells arranged concentrically around a lumen [17]. Acinar cells are highly polarized with the broader base of the pyramidal cells being basophilic and filled with the nucleus and rough endoplasmic reticulum. The narrow apices of acinar cells, facing the lumen contain eosinophilic zymogen granules filled with digestive enzymes [15]. Zymogen granules take up to 30% of the total acinar cell volume, and are released by exocytosis, mediated by multiple G-protein-coupled-receptors in the basolateral plasma membrane. These receptors can recognize a variety of molecules: cholecystokinin (CCK), acetylcholine, gastrin-releasing peptide (GRP), substance P, vasoactive intestinal peptide (VIP) and secretin [18].

The function of the acinar cells is to produce, store and secrete, in a regulated manner, four major categories of digestive enzymes: α-amylases, lipases, nucleases and proteases, responsible for the digestion of carbohydrates, fat, DNA/RNA and proteins, respectively. All these enzymes are first stored in zymogen granules and then secreted into the lumen of the acini by exocytosis. The acinar lumens are connected to small, intercalated ducts that converge into larger intralobular ducts and finally merge into the main pancreatic duct [19]. Once in the ductal system, the digestive enzymes are mixed with water, bicarbonate, Na^+ and K^+, secreted by duct epithelial cells and centroacinar cells. This mixing results in the pancreatic juice [20].

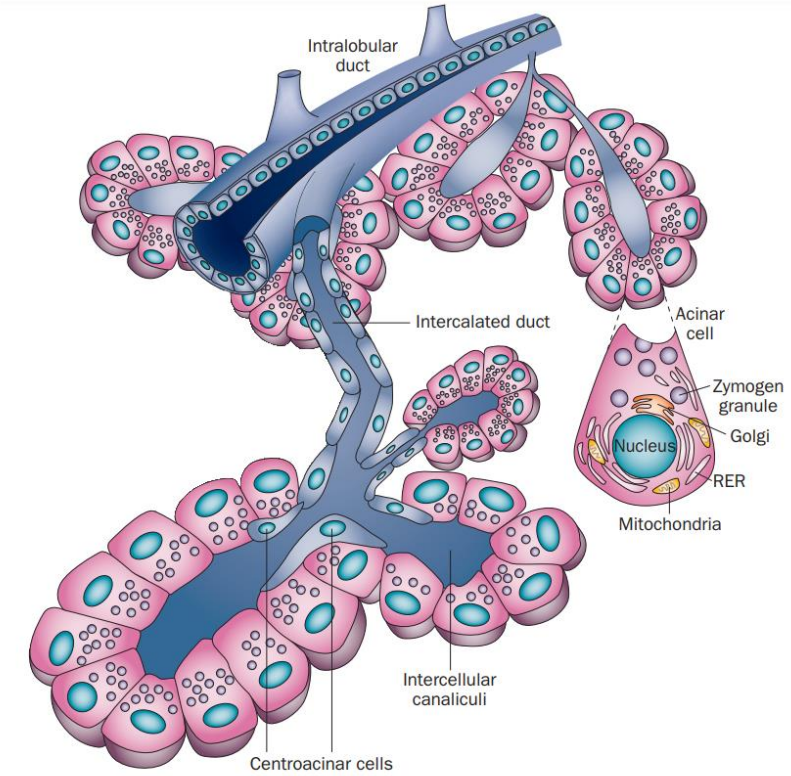

Figure 3. Components of the exocrine pancreas. The exocrine pancreas is comprised of acinar cells and ductal structures. Acinar cells form groups named acinus that surround a central lumen leading into the ductal system. Pancreatic acinar cells produce, store and secrete enzymes necessary for the digestion and absorption of food in the small intestine. Digestive enzymes are produced in the basal region of acinar cells and secreted through the apical membrane of the acinar cell into small intercalated ducts that are directly connected to increasingly larger intralobular ducts that join the main pancreatic duct. Adapted from ref [4]

1.2 Diseases of the pancreas

1.2.1 Diabetes mellitus

Diabetes mellitus (DM) is a severe chronic disease of the endocrine pancreas. DM is characterized by chronically elevated blood sugar caused by impairment in the production of insulin (reduction in the amount produced or total loss of production) or too little effect of the insulin produced (insulin resistance) [21]. DM is the chronic disease with the highest growth rate worldwide [22], with an estimated of 573 million adults aged 20-79 years presenting any form of the disease and an estimated projection of around 783 million patients by 2045 [21]. DM results in increased morbidity and mortality and is, therefore, a burden for the individual as well as a great challenge for the health systems worldwide [23].

DM is classified into many subtypes, with type 1 and type 2 diabetes being the most common. Type 1 diabetes (T1D) is the result of an autoimmune response where the immune system attacks the insulin-producing β-cells of the pancreas. This autoimmune response leads to very reduced or no insulin production [21] and the need for patients to receive exogenous insulin replacement as a treatment [24]. The factors leading to this disease are englobed in two main categories: genetic predisposition and environmental factors (viral infections, gut microbiome, vitamin D, breast feeding, etc) [25].

Type 2 diabetes (T2D) is characterized by insulin resistance on the target organs and β-cell inability to keep up with increased insulin production requirements [26], being this the main cause of hyperglycaemia. This variant includes around 90% of the patients diagnosed with diabetes [21] and the number of cases is rapidly increasing due to the growth in obesity and physical inactivity and modification of dietary patterns [26].

When compared to T1D, T2D has a milder onset and less probability of developing symptoms. Moreover, it is hard to determine when the clinical picture will manifest, therefore the pre diagnosis period tends to be long, leading to mistreatment of nearly 50% of the patients [21]. Furthermore, T2D is an age-related disease, increasing prevalence in the older sectors of the population, which leads to numerous secondary effects such as loss of motor capacity [27], increase in cardiovascular affections [28] or generation of renal deficiencies [29].

Monogenic diabetes is a rare subtype of diabetes, accounting for 1-2 % of all diabetes cases, where, unlike the rest of the subtypes, the disease is caused by a mutation in a single gene [30]. Monogenic diabetes can further be divided into neonatal diabetes, occurring in new-borns and infants, and maturity-onset diabetes of the young (MODY), which takes place in adolescents and young adults. Neonatal diabetes has been linked to more than 20 different genetic causes, and it can manifest as transient, permanent or syndromic forms [31]. MODY, the most common type of monogenic diabetes, is characterized by an autosomal dominant pattern of inheritance and progressive β-cell dysfunction [32, 33]. Onset often presents before 25 years of age. As to this day, there are at least 14 different types of MODY identified [32], with mutations in genes encoding the metabolic enzyme glucokinase (GCK) and hepatocyte nuclear factors (HNFs) being the most common [34].

1.2.2 Pancreatitis

Pancreatitis is one of the leading causes for gastrointestinal disease hospital admissions and is associated with considerable morbidity, mortality and socioeconomic burden [35]. The disease is normally divided into acute pancreatitis, recurrent acute pancreatitis and chronic pancreatitis, which together form a disease continuum [36]. The progression of a sentinel attack of acute pancreatitis to recurrent acute pancreatitis and eventually to chronic pancreatitis is often driven by chronic alcohol consumption or genetic risk factors. Mechanistically, the genes predisposing for chronic pancreatitis can be classified into three aetiologies, the trypsin-dependent, the misfolding-dependent and the ductal pathway [37].

Acute pancreatitis has an incidence of 34 per 100000 people-year in high-income countries [38]. The diagnosis is based on meeting at least two of three criteria: 1) upper abdominal pain, 2) serum amylase or lipase (or both) levels of at least three times the normal upper limits or 3) findings consistent with acute pancreatitis on radiological imaging [39]. Most patients present mild acute pancreatitis which is self-limiting and usually resolves within a week. Around 20% of cases develop moderate or severe acute pancreatitis, with necrosis of the pancreatic or peripancreatic tissue or organ failure and a substantial mortality rate of 20-40% [40]. The most common causes of acute pancreatitis are gallstones (45% of cases) and alcohol abuse (20%) [41], with other factors such as medication, hypercalcemia, hypertriglyceridemia, infection, genetics, autoimmune diseases and trauma playing a minor role [40].

Chronic pancreatitis is a fibroinflammatory syndrome of the pancreas, where inflammatory episodes lead to irreversible pancreatic tissue damage [42, 43]. Hallmarks of chronic pancreatitis are pain, pancreatic fibrosis, parenchymal or intraductal calcifications and exocrine pancreatic insufficiency resulting in malabsorption. Diabetes and increased risk of pancreatic cancer are long-time effects of the disease [44]. Excessive alcohol consumption and regular tobacco use are the risk factors with highest prevalence amongst chronic pancreatitis patients although they rarely lead to development of chronic pancreatitis on their own. Other risk factors are pancreatic duct obstruction, hypertriglyceridemia, chronic kidney disease and IgG4-related disease (resulting in autoimmune pancreatitis) [44]. Importantly, multiple genetic factors can increase the risk of developing chronic pancreatitis. Variants of serine peptidase inhibitor kazal type 1 (*SPINK1*), chymotrypsin C (*CTRC*), chymotrypsinogen B1-chymotrypsinogen B2 (*CTRB1-CTRB2*), cystic fibrosis transmembrane conductance regulator (*CFTR*), claudin 2 (*CLDN2*) or MORC family CW-type zinc finger protein 4 (*MORC4*) are among the genetic risk factors identified [44].

Around 1% of patients with chronic pancreatitis present a hereditary form. Hereditary pancreatitis (HP) is an autosomal dominant disease with a penetrance around 80%. Although, in the clinical setting sometimes it is not possible to determine the pattern of inheritance. HP is characterized by an early onset around 5-10 years of age [45, 46]. HP symptoms include repeated attacks of severe abdominal pain, decreased endocrine and exocrine pancreatic function, nausea, vomiting, diabetes, pseudocysts, bile duct and duodenal obstruction and pancreatic cancer [47]. The disease is mainly associated with mutations in the cationic trypsinogen gene (PRSS1). The most common disease-associated mutations are R122H and N29I. These mutations lead to increased trypsin stability and trypsinogen autoactivation respectively [45, 47].

1.2.3 Pancreatic cancer

Malignant disease in the exocrine pancreas, commonly referred to as pancreatic cancer (PC), is a highly invasive type of cancer, where most patients do not show obvious symptoms during the development of the disease. This makes early diagnosis difficult, leading to advanced stages of the disease when detected [48]. PC accounts for around 2% of all cancers and about 5% of cancer-related deaths worldwide, being the fourth cause of cancer death in western societies [49]. PC incidence has increased over the past few years due to ageing of the world's

population, as well as increases in the prevalence of modifiable pancreatic risk factors such as diabetes, cigarette smoking and obesity [50].

Risk factors for PC can be classified as non-modifiable and modifiable. Non-modifiable risk factors are age, sex, area, blood group, family history and genetic susceptibility, while modifiable risk factors include intestinal microflora, smoking, alcohol consumption, chronic pancreatitis, obesity, dietary factors and infections [51]. From this list, cigarette smoking, chronic pancreatitis, age and family history of the disease are the ones with a greater effect [52].

PC can be divided into endocrine pancreas tumours and non-endocrine pancreas tumours. Endocrine pancreas tumours are uncommon with pancreatic neuroendocrine tumours (PNETs) being the most common ones. PNETs emerge from the endocrine cells of the pancreas and can result in the depletion of hormonal production or excess production of hormones [53]. Exocrine pancreas tumours are more common where the most common pancreatic neoplasm is pancreatic ductal adenocarcinoma (PDAC), accounting for more than 90% of pancreatic cancer cases [54]. PDAC develops from intraductal papillary mucinous neoplasms or mucinous cystic neoplasms, and from non-invasive precursor lesions like pancreatic intraepithelial neoplasia [55]. Characteristic somatic mutations are normally found in PDAC development. The most common is activation of the *KRAS* oncogene (around 90%), followed by inactivation of the tumour-suppressor genes including *TP53*, *CDKN2A* and *SMAD4* [56].

1.3 Carboxyl ester lipase (CEL)

Carboxyl ester lipase (CEL) is an enzyme produced by the acinar cells of the pancreas. It is one of four major lipases secreted by the pancreas into the duodenum, being the only one that is stimulated by bile salts [57-59]. For this, the protein is also referred to as bile salt-dependent lipase (BSDL) [60] or bile salt-stimulated lipase (BSSL) [61]. CEL exhibits wide substrate specificity [62, 63] and once activated it can hydrolyse dietary fat, cholesteryl esters and fat-soluble vitamins [57, 58, 63, 64]. CEL is estimated to represent around 4 % of the total proteins detected in the pancreatic juice [64]. The enzyme is also expressed in the lactating mammary gland, where it comprises 1-2 % of total proteins secreted in the mother's milk [57, 65].

1.3.1 The human *CEL* locus

The human *CEL* gene is situated in chromosome band 9q34.13 (**Figure 4**), where it covers around 10 kilobase pairs [66, 67]. The gene is composed of 11 exons, with exon 11 containing a GC-rich region of repeated segments, each of 33 bp. Such an arrangement of almost identical nucleotide sequences is denoted a VNTR (variable number of tandem repeats). At the protein level, the VNTR codes for repeats of 11 amino acids. The VNTR region makes *CEL* highly polymorphic, as multiple VNTR lengths have been described, ranging between 3 and 23 repeats for the human gene. However, 16 repeats is the allele with the highest frequency in the human population [68, 69].

Located in the same locus as *CEL*, but eleven kb downstream of the gene, is a *CEL* pseudogene (*CELP*) [67, 70]. *CELP* lacks exons 2-7. Except from that, *CEL* and the pseudogene present highly similar genomic sequences with a 97% identity [70]. *CELP* is not expected to be translated into any functional protein as it contains a stop codon in its second exon, resulting in a truncated protein [71]. It has been proposed that *CEL* is a duplicated gene of *CELP* that retained its activity, while *CELP* was the original gene that became inactivated [70].

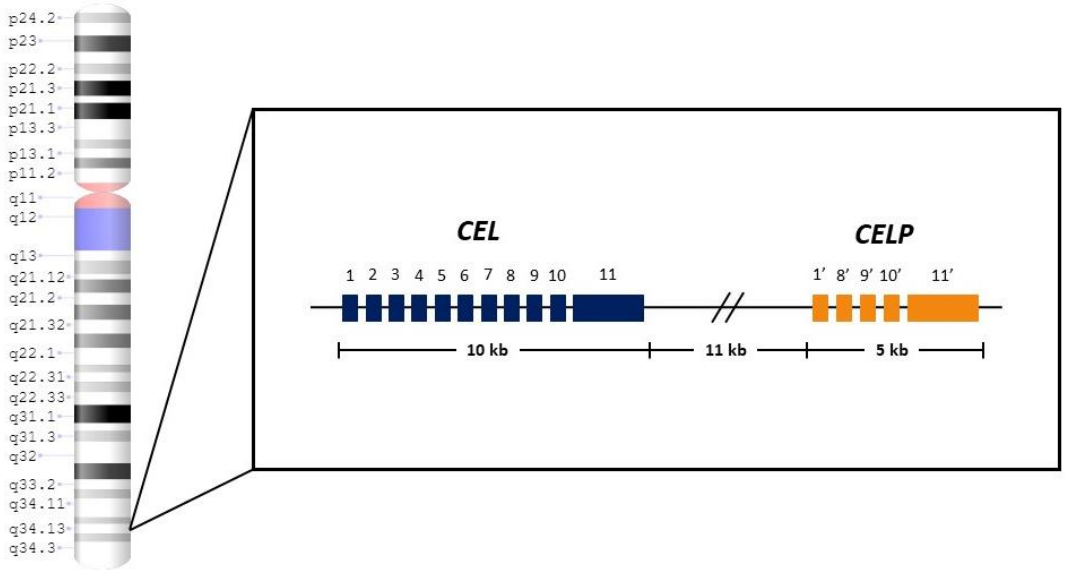

Figure 4. Schematic structure of the human *CEL* locus. The locus is positioned on chromosomal band 9q34.13. The *CEL* gene (dark blue) consists of 11 exons covering 10 kb. Eleven kb downstream from *CEL*, the pseudogene *CELP* is located (orange), having only 5 exons, corresponding to exons 1,8,9,10 and 11 of *CEL*. Adapted from ref. [80]

1.3.2 The CEL protein

The structure of CEL is divided into two different parts. The first part of CEL is the N-terminal globular domain, a catalytic domain containing the catalytic site (Ser-194, His-435 and Asp-320), multiple binding sites for bile salts, and a signal peptide [72, 73]. The second part of CEL is the C-terminal sequence, containing a VNTR encoded domain [74] resulting in an intrinsically disordered protein tail. This VNTR encoded region is enriched in proline (P), glutamic acid (E), serine (S) and threonine (T) known as the PEST sequence [75] (**Figure 5**).

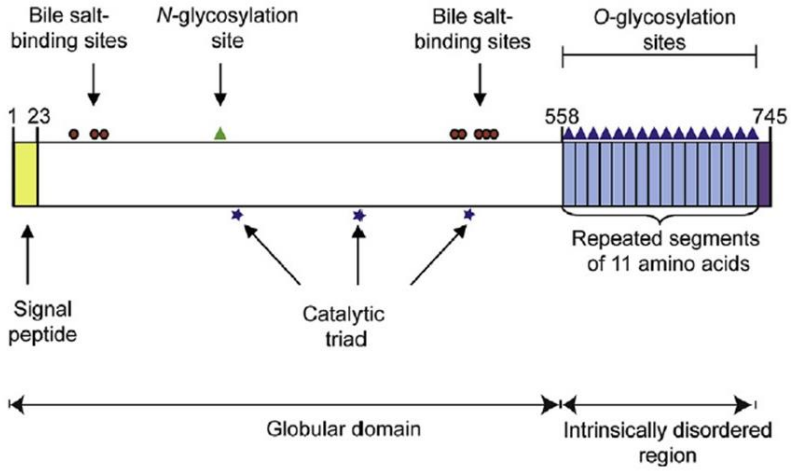

Figure 5. Carboxyl ester lipase (CEL) schematic protein structure. CEL contains two different structural domains: a globular N-terminal domain responsible for the catalytic activity and an intrinsically disordered region in the C-terminal. N- and O-glycosylation sites, signal peptide, bile salt-binding sites and the catalytic triad are indicated by arrows. This figure refers to the most common variant of CEL with a VNTR of 16 repeats. Figure from ref. [72]

CEL is a glycoprotein that contains two types of glycans attached, N-linked and O-linked glycans. The N-glycan is covalently linked to the residue Asn187 in the globular domain during the translation in the endoplasmic reticulum (ER) with the aid of glucose-regulated protein 94 (GRP94) [76]. This N-glycosylation plays a role in the correct folding and secretion of CEL. Additionally, multiple O-glycans are placed onto the serine and threonine residues of the PEST sequences in the Golgi apparatus [77]. This O-glycosylation has been suggested to play a major role in masking of the PEST sequences, in order to avoid protein degradation [75]. Once glycosylated, CEL gets phosphorylated at Thr340 allowing translocation through the secretory pathway [78], and the mature version of CEL gets stored in zymogen granules together with other digestive enzymes. The most common CEL variant, encoded by a VNTR of 16 repeats, has a predicted mass of approximately 79 kDa However, due to the post-translational modifications mentioned above, the mature protein with 16 repeats can weigh up to 120 kDa [79].

1.4 Disease-associated and unknown pathogenicity variants of CEL

CEL is a highly variable protein, especially in the VNTR encoded region where most of the pathogenic and non-pathogenic variants present their variability. Apart from VNTR length variation, multiple variants have been described so far, ranging from pathogenic variants (CEL-MODY) [68], risk variants for chronic pancreatitis (CEL-HYB1) [80] to variants of unknown pathogenicity (CEL-3R, CEL-INS) [69, 72].

1.4.1 Insertion variants

Insertions of one additional cytosine in the poly-C tract of some of the VNTR repeats have been observed, resulting in premature stop-codons and truncated CEL proteins. Insertions were described in repeats 4, 7, 8, 9, 10 and 11 and have been associated with faecal elastase deficiency in diabetic patients [72]. The combined frequency of single-bp insertions in normal controls of Northern European descents is around 0.07, therefore the elimination of the unique C-terminal amino acid sequence (KEAQMPAVIRF) does not have a profound biological impact [68].

These single base pair insertions result in the formation of a new C-terminal amino acid sequence (PRAAHG), differing from the normal sequence (PAVIRF). This newly generated sequence has been proposed as a diagnostic tool, as CEL insertions were suggested to emerge as early somatic mutations in PDAC [81].

1.4.2 The CEL-3R Variant

A rare variant of CEL denoted as CEL-3R, only containing 3 repeats of the VNTR encoded region, was described in a Danish family in 2010. Seven carriers were found when analysing the family. Four of them presented diabetes, one showed impaired fasting glycemia and one had impaired glucose tolerance. Demonstrating the possibility of this short variant predisposing to diabetes. Nevertheless, this co-segregation may have been caused fortuitously. Moreover, there was no evidence of exocrine deficiency. Finally, contrasting to other CEL variants, CEL 3-R results in a short, but normal C-terminal tail [69].

1.4.3 The CEL-MODY Variant

The disease CEL-MODY, also denoted MODY8, is caused by mutations in the CEL VNTR that lead to a syndrome with both an exocrine and endocrine pancreatic phenotype. This inherited disease is clinically characterized by development of maturity-onset inherited diabetes , slowly progressing pancreatic exocrine dysfunction, fatty replacement of pancreatic parenchyma and development of pancreatic cysts [68, 69, 82].

CEL-MODY was originally discovered in two Norwegian families [68, 69] caused by single-base-pair deletions located in the first (DEL1, c.1686delT) and fourth (DEL4, c.1785delC) repeat of the CEL VNTR. Recently, one patient in Italy (c.1818delC) [83] presented a deletion in the fifth repeat of the CEL VNTR, which was denoted as a new pathogenic variant leading to CEL-MODY. Moreover, two more families (one Swedish and one Czech) [84] were identified with variants leading to CEL-MODY, where the Swedish family presented a deletion in the first repeat (c.1685delC) while the Czech family presented a deletion in the fourth repeat of the CEL VNTR (c.1786delG). Both of these variants presented differences at the DNA level to those presented by Ræder *et al* [68], however, at the protein level both variants were predicted to have similar consequences to DEL1 or DEL4 variants respectively [84].

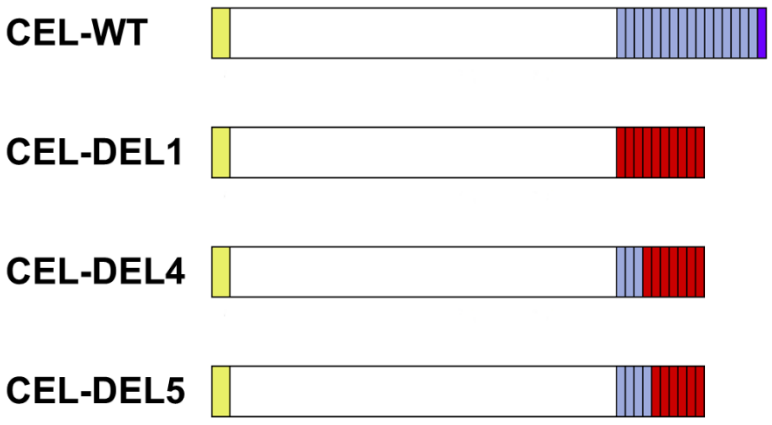

Figure 6. Schematic overview of the different CEL-MODY variants discovered. CEL-WT is represented with 16 repeated segments, the most common number of repeats in the human. CEL-DEL1 corresponds to one of the original Norwegian families (c.1686delT) and the Swedish family (c.1685delC). CEL-DEL4 correlates with the second Norwegian family (c.1785delC) and the Czech family (c.1786delG). CEL-DEL5 represents an isolated case in Italy (c.1818delC). CEL-DEL5 is not recognized as an official name for this variant. Elements are not drawn to scale.

These single-base-pair deletions lead to a shift of the reading frame, which results in the translational machinery encountering an early stop codon and producing truncated CEL protein. The aberrant proteins present a shorter C-terminal formed of 11-amino acid repeats, with an altered amino acid sequence compared to the wildtype, which leads to altered biochemical properties and a reduced number of potential O-glycosylation sites [68]. Moreover, functional studies of DEL1 in cell lines have shown a tendency to form both intra- and extracellular aggregates, which differs from the behaviour of the wildtype protein [82]. This has been demonstrated to cause an impairment in the secretion of CEL, which remains inside the cell forming insoluble aggregates than can induce ER-stress and apoptosis [85, 86].

Interestingly for this thesis, the mutation leading to CEL-MODY introduces cysteine residues in the VNTR region, these residues have been hypothesized to form disulphide bridges which may lead to an increase in protein aggregation and therefore the increase in ER-stress, activation of the unfolded protein response and apoptosis [85, 86].

1.4.4 The CEL-HYB Variant

CEL-HYB1 is a hybrid allele discovered and described as pathogenic in 2015. The allele was described as a genetic risk factor for chronic pancreatitis, being overrepresented by five-fold in chronic pancreatitis patients compared to healthy controls [80]. This hybrid variant is probably the result of a non-allelic homologous recombination, where a crossover between intron 10 of *CEL* and the neighbouring pseudogene *CELP* (**Figure 7**), gave place to the CEL-HYB variant containing exons 1-10 from *CEL* gene and exon 11 of the pseudogene *CELP* [80]. This may have been caused by the high sequence similarities between *CEL* and *CELP* [87].

The *CEL-HYB1* allele leads to generation of a chimeric enzyme, composed of intact CEL globular domain and a short VNTR with a length of only 3 repeats, caused by the stop codon present in *CELP* VNTR. This chimeric protein presents a lower enzymatic activity than the normal CEL-WT, as well as reduced secretion and intracellular retention [80]. Similarly to CEL-MODY, this variant has also been shown to induce autophagy and ER stress, likely belonging to the misfolding-dependent pathway of genetic risk in chronic pancreatitis [88].

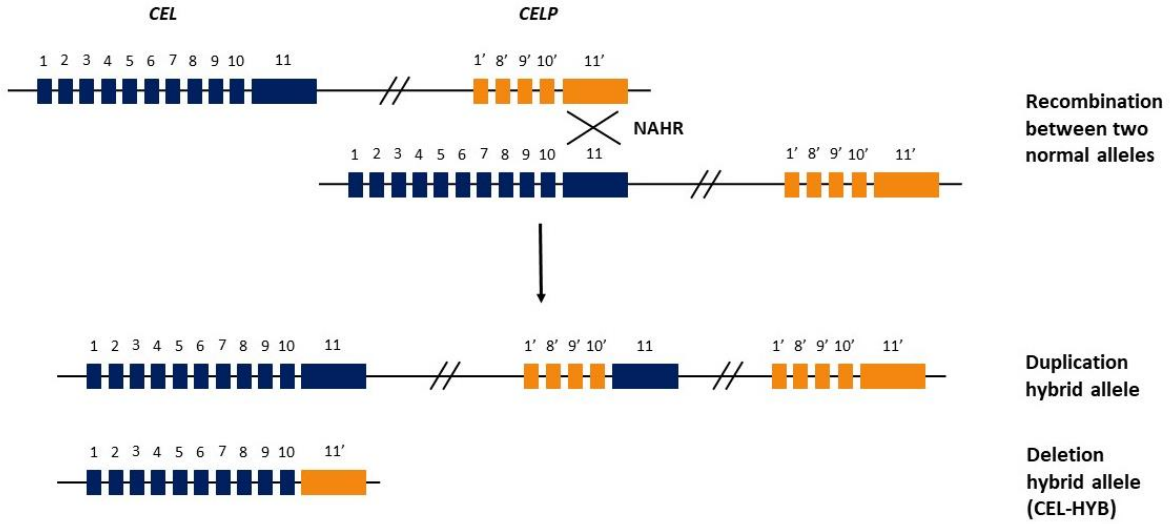

Figure 7. Schematic structure of the duplication hybrid allele and the deletion hybrid allele (*CEL-HYB*) and proposed mechanism for the generation of *CEL-HYB*. Probable mechanism for the generation of *CEL-HYB* by non-allelic homologous recombination (NAHR) between *CEL* and *CELP*. With the recombination occurring in the region between exons 10 and 11. Adapted from ref. [80].

1.5 Protein folding and misfolding

In order to be biologically active, proteins need to fold into their native three-dimensional structure. The process by which proteins achieve their native structure is defined as protein folding [89]. Protein folding is a complex process that requires multiple macromolecules and depends on the environment in which folding takes place. Polypeptides can fold in the cytoplasm after release from the ribosome or in other subcellular compartments such as the endoplasmic reticulum [90, 91]. *In vivo*, proteins encounter challenges regarding folding, caused by the space being crowded with macromolecules. This can lead to incompletely folded peptide chains exposed, which can aggregate with other molecules. To avoid this, the cellular environment has evolved to prevent protein misfolding and aggregation. Some examples of these defence mechanisms are chaperones and the ubiquitin-proteasome system [91].

1.5.1 Protein folding in the endoplasmic reticulum

The endoplasmic reticulum (ER) is a specialized compartment of the eukaryotic cells, where synthesis, folding and transport of at least one-third of the proteins in the eukaryotic cell occur [92]. The ER regulates folding, assembly, trafficking and degradation of proteins destined for organelles (ER, Golgi apparatus, lysosomes, and plasma membrane) and the extracellular space. The dynamic life cycle of a protein begins in the ribosome. Newly generated

polypeptides are translocated into the ER through the translocon complex thanks to their N-terminal signal [93]. Once in the ER, the signal sequence is removed and the polypeptides are modified, assembled and folded to acquire their native conformation [92]. Maturation of nascent polypeptide chains is vulnerable to misfolding and aggregation. As a result, the ER contains molecular chaperones, a set of proteins that aid other proteins acquire their native active conformation [94]. Furthermore, the folding process is enhanced by different ER resident enzymes. For example, protein disulphide isomerases (PDIs) and peptidyl prolyl isomerases increase the rate and efficiency of disulphide bond formation and *cis-trans* isomerization at proline residues respectively [95, 96]. Moreover, the ER facilitates proper folding due to its oxidizing nature and Ca^{2+} concentration, facilitating disulphide bond formation and proper chaperone function. Once proteins are correctly assembled, they leave the ER and continue through the secretory pathway [92].

1.5.2 Protein misfolding and endoplasmic reticulum stress

High-quality protein folding is essential for cell survival, function, and normal organismal physiology. However, during the folding process newly formed polypeptides are prone to misfolding or aggregation. Altered homeostasis in the ER leads to accumulation of misfolded or unfolded proteins in the ER lumen, known as ER stress [92]. As a counter measure, the ER presents control mechanisms to ensure correct folding, modification, assembly and transportation of proteins. To cope with the stress, cells activate the unfolded protein response (UPR) (**Figure 8**). The UPR is an integrated intracellular signalling pathway that communicates with the cytoplasm and the nucleus to transmit information about the protein folding status inside the ER [97]. Under acute stress, the UPR is activated to restore protein-folding homeostasis. For this, the UPR transiently attenuates protein synthesis, increases folding and transport in the ER, and upregulates protein degradation and autophagy [92]. The UPR depends on three key signal activators, PRKR-like ER kinase (PERK), inositol-requiring protein 1α (IRE1α) and activating transcription factor 6α (ATF6α), that give rise to separate branches of the response. Together they induce a set of transcriptional and translational events to restore ER proteostasis [98-100]. However, if ER stress is too severe and the protein-folding defect is not corrected, proteotoxicity can lead to cell death promotion, derived in apoptosis [92].

In vivo, ER stress is induced by both intrinsic (protein synthesis increase, gene mutations leading to misfolding and defects in UPR pathways) and extrinsic factors (inflammation, hypoxia or nutrient deprivation). As such, ER stress has been shown to play a crucial role in immune regulation, antitumor response, disease progression and complex chronic inflammation. Given that all these factors are involved in pathogenesis, it is not unexpected that ER stress contributes to the start and development of many diseases [92]. Consequently, efects in the UPR are emerging as key contributors to diseases such as diabetes, metabolic diseases, atherosclerosis, fibrosis, renal disease, neurodegeneration, and cancer [92, 101, 102].

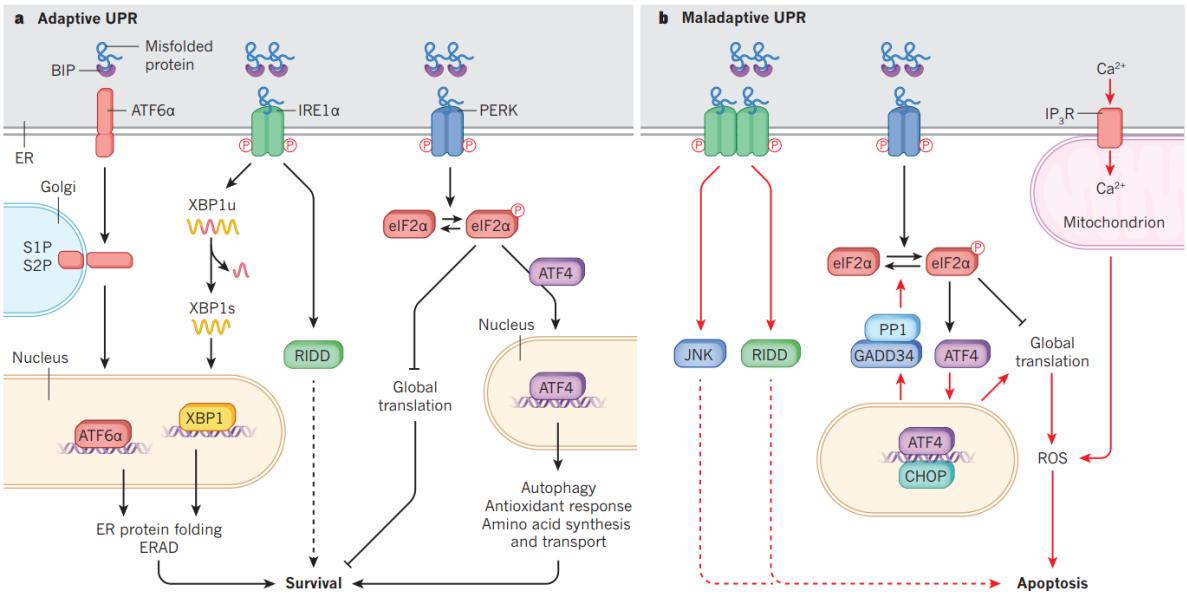

Figure 8. Unfolded protein response schema. A) In the presence of unfolded or misfolded proteins BiP interacts with them and is detached from the ER-luminal domains of ATF6α, IRE1α and PERK. This activates the UPR adaptative pathway in order to restore ER homeostasis B) Maladaptive UPR is induced by sustained activation of the PERK pathway, as a result of prolonged severe ER stress. Activation of this pathway leads to apoptosis. Figure from ref. [92]

1.5.3 Intrinsically disordered proteins and protein disordered regions

Although most proteins require a specific native three-dimensional structure to be active, some proteins do not present a fixed structure. Intrinsically disordered proteins (IDPs) are proteins with entirely disordered sequences, unable to adopt any tertiary structure with well-organized hydrophobic core, while intrinsically disordered protein regions (IDRs) are polypeptide segments that do not present a fixed tertiary structure, although they exist in a heterogenous ensemble of conformations, but are embedded into proteins which adopt a well-defined structure [103]. IDPs and IDRs typically exhibit low sequence complexity and a bias in amino acid composition. Thus, they tend to be enriched in polar (arginine, glycine, glutamine, serine,

proline, glutamine and lysine) and structure-breaking (glycine and proline) amino acid residues [104, 105]. IDPs and IDRs are highly flexible, and considered to undergo transitions to more ordered states or fold into stable secondary or tertiary structures when binding to their targets, although these properties are dependent on sequence and environment [106].

IDRs have been shown to play a role in diverse functions that are mediated by proteins and that depend on the amino acid sequence. For example, polar and charged IDRs may increase solubility, improving expression and purification in vitro [107]. Other IDRs can confer the ability of aggregation [108]. Moreover, IDRs with repeated sequence elements, especially those containing disorder-promoting amino acids such as those mentioned above, play a role in the formation of biomolecular condensates through phase separation, as they provide the basis for multivalent weakly adhesive intermolecular interactions [109, 110]. Furthermore, IDRs are prone to contain post-translational modification sites due to their accessibility for modifying enzymes, offering even more mechanisms for functional regulation or the regulation of phase behaviour [111-113]. The sequence variability, structural flexibility and post-translational modification possibilities of IDRs make them perfect candidates for elements of multifunctional proteins. Here, they provide what is known as multispecificity, i.e. the property of interacting with many partners [106]. Some examples of this promiscuity are moonlighting proteins, proteins in which one polypeptide chain exhibits more than one physiologically relevant function. Moonlighting proteins are able to adopt dramatically different structures in different complexes [114]. In summary, amino acid sequences that constitute IDRs encode a wide range of chemical properties and provide a complex and chemically heterogenous toolkit for biomolecular assembly and biological functions [103].

IDRs normally exist to serve a specific biological role. However, variations in expression levels of the corresponding proteins, covalent modifications, environmental changes or gene mutations may lead to an altered role or malfunctioning. Accordingly, IDRs are commonly found in proteins involved in human disorders including cancer, cardiovascular diseases, amyloidosis, neurodegenerative diseases and diabetes [106, 115]. For example, IDRs represent integral components of pathogenesis in neurodegenerative diseases, ranging from the relatively common disease Alzheimer’s and Parkinson’s, to the rare inherited conditions Charcot-Marie-Tooth, Huntington’s and spinal muscular atrophy [116]. Furthermore, prion diseases are also characterized by the presence of IDRs where the prion protein presents

unstructured regions that play a certain role in modulating and propagating aggregation [117, 118]. Finally, IDPs and IDRs have been associated with diabetes in different ways. For example, amylin mutations or amyloid fibril formation have been observed to be a common feature of diabetes type 2 [119-121] and mutations in the *CEL* causes the MODY8 syndrome (see **Chapter 1.4.3**), characterized by exocrine and endocrine dysfunction of the pancreas.

1.6 Cysteine residues and disulphide bonds in protein folding

Cysteine residues play essential roles in protein structure and function. They are considered as order promoting amino acids, being present in highly conserved motifs, providing a high degree of structural order and stability. This is achieved through disulphide bond formation, which influences the folding and structure of proteins, and therefore can have a fundamental impact on the function of a protein [109], on the maintenance of proper maturation and on localization through protein-protein intermolecular interactions [122]. Disulphide bonds form in oxidative environments and can be dissociated under reductive conditions. This oxidative environment is normally found inside the ER, where around 30% of the proteins require disulphide bond formation to be properly assembled [123]. However, oxidative protein folding is probably the most complicated protein folding scenario, as the possibility of protein misfolding increases with the increasing number of cysteines in the protein, where only one disulphide bond pattern is proper for the correct folding [124]. To facilitate this complex folding process, disulphide bond formation inside the ER is mediated by protein disulphide isomerases (PDIs). PDIs are a set of enzymes that have a critical role in oxidative protein folding, as they catalyse disulphide bridge formation, isomerization or reduction steps depending on the conditions [125]. The oxidoreductase capacity of the PDIs is essential to oxidative folding, as both the formation of native disulphide bridges and reduction of non-native erroneous disulphide bonds is essential towards proper folding. Although it is still unknown how the PDIs differentiate between native and non-native disulphide bridges, some theories revolve around the idea of native disulphide bridges burial inside the folded proteins, leaving only non-native bridges exposed for PDI interaction [125]. Based on the importance of oxidative folding, it is not unusual that cysteine residues and disulphide bridges are central to many diseases, either by changes in the number of cysteines or changes in the protein environment, normally affecting a protein's activity, folding and aggregation.

Diseases caused by mutations affecting cysteines can range from those caused by removal of cysteines, inclusion of extra cysteines or even some that can be caused by either removal or introduction of cysteines. One example of a disease where substitution of cysteines for another amino acids could play a role, is chronic pancreatitis, caused by carboxypeptidase A1 (CPA1). CPA1 presents a mutation on a disulphide-stabilized loop (p.C271R) that causes loss in protein secretion. Moreover, this was supported by artificial mutations of this cysteine (p.C271A) and its binding partner (p.C258A) for alanine, which lead to abolishment of CPA1 secretion. Furthermore, mutations inside this loop introducing amino acids with long side chains (Lys, Arg, Met) also caused a reduction in secretion, probably due to abolishment of the disulphide bond [126].

As mentioned, substitution of different amino acids with cysteines can also lead to disease. This can be exemplified by the rare connective tissue diseases thanatophoric dysplasia and achondroplasia. Multiple pathogenic cysteine mutations have been described for fibroblast growth factor receptor 3 (FGFR3), with two of them related to thanatophoric dysplasia (p.G370C, p.S371C) and the other one related to achondroplasia (p.G375C). These mutations lead to a higher degree of dimerization of the transmembrane domains, leading to unregulated signalling and development of the already mentioned pathologies [127]. Finally, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy, a common hereditary form of stroke, can be caused by mutations in Notch3 where cysteine residues are either being mutated into other amino acids or other amino acids are being transformed into cysteines, these leaves Notch3 with an odd number of cysteines and an unpaired cysteine residue, which affects its folding and function [128-130].

In conclusion, due to the importance of structural and regulatory disulphide bonds in membrane and secretory proteins, mutations in cysteines can have dramatic consequences [131]. Acquisition or loss of cysteines often causes retention of mutated proteins in the ER by thiol-mediated mechanisms [132] and aberrant thiol-mediated interactions via unpaired cysteines can directly cause ER retention and aggregation besides misfolding and lead to severe diseases [131].

2. Aims

CEL is an extremely polymorphic gene with some variants causing pancreatic disease. Both types of verified pathogenic mutations (*CEL-MODY* and *CEL-HYB1*) affect the C-terminal region of the CEL protein and involve the generation of cysteine residues in the repeated tail segments. The overall aim of this study was to understand the role of cysteines in the tail regions of the CEL-MODY and CEL-HYB1 proteins variants.

The specific aims were:

1) To analyse how the cysteine residues in CEL-MODY and CEL-HYB1 affect solubility and localization inside the cell.
2) To determine if the cysteines in CEL-MODY and CEL-HYB1 influence endoplasmic reticulum stress in cells expressing these variants.
3) To test whether the cysteines have an effect on O-glycosylation of CEL-MODY.

3. Materials

Table 3.1 Plasmid constructs

Plasmid	Encoding	Description
pcDNA 3.1 CEL-WT/V5-His	CEL-WT	Plasmid expressing CEL wild type (WT) protein containing a 16 repeats VNTR region
pcDNA 3.1 CEL-MODY-CYS/V5-His	CEL-MODY-CYS	Plasmid expressing CEL MODY variant leading to MODY8
pcDNA 3.1 CEL-MODY-ALA/V5-His	CEL-MODY-ALA	Plasmid expressing CEL MODY with cysteines substituted for alanines
pcDNA 3.1 CEL-HYB1/V5-His	CEL-HYB1	Plasmid expressing CEL-HYB1 protein
pcDNA 3.1 CEL- HYB1 C583A/V5-His	CEL-HYB1-C583A	Plasmid expressing CEL-HYB1 protein with Cys 583 mutated to Ala
pcDNA 3.1 CEL- HYB1 C587A /V5-His	CEL-HYB1-C587A	Plasmid expressing CEL-HYB1 protein with Cys 585 mutated to Ala
pcDNA 3.1 CEL- HYB1 C583A/C587A /V5-His	CEL-HYB1-C583A/C587A	Plasmid expressing CEL-HYB1 protein with Cys 583 and 587 mutated to Ala

*All plasmids are based on the mammalian expression vector pcDNA 3.1/V5-His B from Invitrogen. All CEL variants contain a V5 and HIS Tag at the C-terminal. Sequences presented in Appendix A

Table 3.2 Plasmid purification and agarose gel electrophoresis

Product	Catalog number	Supplier
ImMedia™ Amp Agar	45-0034	Invitrogen
LB Broth (Lennox)	L7275-500TAB	Sigma-Aldrich
Ampicillin sodium salt	A9518-5G	Sigma-Aldrich
TE buffer, pH 8	1018499	QIAGEN
QIAfilter Plasmid Maxi Kit	12262	QIAGEN
NuSieve GTG (agarose)	50084	LONZA
TBE Buffer x10	A3945, 1000	PanReac, AppliChem
Ethidium Bromide (0.625 µg/µL)	E406-15ml	VWR
6x DNA gel loading buffer	G2526	Sigma-Aldrich
1 kb DNA Ladder	N3232S	New England Biolabs

Table 3.3 Sanger Sequencing

Product	Catalog number	Supplier
BigDye Terminator v.3.1 Cycle sequencing kit	4337455	Thermo Fisher Scientific
Big Dye terminator v.3.1 sequencing buffer	4336697	Thermo Fisher Scientific
Sephadex ® G-50 Superfine	G5050-50G	Sigma-Aldrich
Betaine	B0300	Sigma-Aldrich

*Primers are described in Table 4.2

Table 3.4 Cell Culture

Product	Catalog Number	Supplier
HEK 293 cells	MCRL-1573	ATCC
Simple Cells		Steentoft *et al.* [133]
Dulbecco's Phosphate Buffered Saline (DPBS)	14190-144	Gibco
DMEM (1X) + GlutaMAX-I	31966-021	Gibco
Fetal Bovine Serum (FBS), Qualified	10270-106	Gibco
Trypsin-EDTA (0.05%), phenol red	25300-054	Gibco
Antibiotic Antimycotic	15240062	Thermo Fischer Scientific
Dimethyl sulfoxide (DMSO)	BP231-1	Fisher bioreagents

Table 3.5 Cell transfection

Product	Catalog Number	Supplier
Lipofectamine™ 3000 Transfection Reagent	L3000008	Invitrogen
Opti-MEM™ (1X) Reduced Serum Medium	31985062	Gibco
Cell culture plate, 6 well, surface: Standard, flat base	83.3920	Sarstedt
Falcon® 12-well Clear Flat Bottom plate	353043	Corning

Table 3.6 Cell lysis and determination of protein concentration

Product	Catalog Number	Supplier
Pierce™ RIPA Buffer	89901	Thermo Fischer Scientific
cOmplete Mini, EDTA-free protease inhibitor cocktail	11836170001	Sigma-Aldrich
Pierce BCA Protein Assay kit	23225	Thermo Fischer Scientific

Table 3.7 SDS-PAGE and Western Blotting

Buffers and solutions	Catalog Number	Supplier
NuPage™ MOPS SDS buffer (20x)	NP0001	Invitrogen
NuPage™ 10%, Bis-Tris Gel, 1.5 mm, 10-well	NP0315BOX	Invitrogen
Bolt™ 10% Bis-Tris Plus, 1.0 mm, 12-well	NW00102BOX	Invitrogen
NuPAGE ® LDS sample buffer (4x)	NP0007	Invitrogen
NuPage ® Sample Reducing agent (10x)	NP0009	Invitrogen
Precision Plus ProteinTM dual color standard	161-0374	BioRad
Trans-Blot Turbo Transfer Pack Mini format, 0.2 µm PVDF	1704156	BioRad
Trans-Blot Turbo Transfer Pack Midi format, 0.2 µm PVDF	1704157	BioRad
Tween®20	P2287	Sigma-Aldrich
SuperSignal™ West Femto Maximum Sensitivity Substrate	34096	Thermo Fischer Scientific

Table 3.7 Immunocytochemistry

Product	Catalog Number	Supplier
ECM Gel	E1270-1ML	MERC
Fibronectin human plasma	F0895-2MG	MERC
16% Formaldehyde Solution (w/v), Methanol-free	28906	Thermo Fischer Scientific
Phosphate-Buffered Saline (PBS) tablets	18912-014	Gibco
Tween20 ®	P2287-100ML	Sigma-Aldrich
Triton™X-100	X100-100ML	Sigma-Aldrich
Glycine	G7126-1KG	Sigma-Aldrich
VWR Microscope Slides	631-1553	VWR
Normal Goat Serum Control	10000C	Invitrogen
Hoechst 33258 solution	94403	MERC
ProLong™ Glass Antifade Mountant	P36982	Thermo Fisher Scientific

Table 3.8 Antibodies

Antibody	Catalogue Number	Supplier	Usage	Dilution
V5 Tag mouse monoclonal antibody	46-0705	Invitrogen	Primary (WB, ICC)	1:20000 1:300
GAPDH (0411) mouse monoclonal antibody	sc-47724	Santa Cruz Technologies	Primary (WB)	1:1000
BiP rabbit polyclonal antibody	3183S	Cell Signaling Technology	Primary (WB)	1:500
PERK (D11A8) rabbit monoclonal antibody	5683S	Cell Signaling Technology	Primary (WB)	1:500
Calnexin (C5C9) rabbit monoclonal antibody	2679S	Cell Signaling Technology	Primary (WB, ICC)	1:1000 1:80
GRP78 BiP rabbit polyclonal antibody	ab21685	Abcam	Primary (ICC)	1:500
GALNT2 rabbit polyclonal antibody	HPA011222	Sigma-Aldrich	Primary (ICC)	1:200
HRP-Goat anti-mouse IgG	626520	Invitrogen	Secondary (WB)	1:10000
HRP-Goat anti-rabbit IgG	656120	Invitrogen	Secondary (WB)	1:10000
IgG (H+L) Cross-Adsorbed F(ab′)2-Goat anti-Mouse, Alexa Fluor™ 488	10328172	Invitrogen	Secondary (ICC)	1:500
IgG (H+L) Cross-Adsorbed F(ab′)2-Goat anti-Rabbit, Alexa Fluor™ 594	10514963	Invitrogen	Secondary (ICC)	1:500

*WB = Western blot, ICC = Immunocytochemistry

Table 3.9 Buffers and solutions

Buffers and solutions	Use	Preparation
NuPage™ MOPS SDS buffer (1x)	SDS-PAGE	For 1 L: 50 ml NuPAGE® MOPS buffer (20x) in 950 MilliQ dH_2O
PBS-Tween 0.1% (PBS-T)	WB ICC	For 1L: 2 PBS tablets and 1 ml Tween®20 and 1000 ml MilliQ dH_2O
5% BSA	WB	5-gram Bovine Serum albumin per 100 mL of 0.1% PBS-T
3% BSA	WB	3-gram Bovine Serum albumin per 100 mL of 0.1% PBS-T
0.2 M Phosphate buffer	ICC	29mL of 0.2 M NaH2PO4 and 72 mL of 0.2 M Na2HPO4. Adjusted to pH 7.2
4% Formaldehyde	ICC	1 mL 16% formaldehyde solution and 1 mL MiliQ dH_2O and 2 mL 0.2 M phosphate buffer
1M Glycine	ICC	15.014 grams of glycine in 200 ml MilliQ dH_2O
100mM Glycine	ICC	20 mL 1M Glycine and 180 mL MilliQ dH_2O
50 mM NH_4Cl	ICC	1.337 grams of NH_4Cl and 500 mL of MiliQ dH_2O
PBS-Triton X-100 0.1%	ICC	For 50 mL: 50 mL PBS and 50 µL Triton X-100
Blocking buffer (5% Goat serum)	ICC	0.5 mL Goat Serum in 9.5 mL PBS

***** WB = Western blot, ICC = Immunocytochemistry

Table 3.10 Technical equipment

Instrument	Manufacturer
Applied Biosystems Thermal Cycler 2720	Thermo Fisher Scientific
3500xL genetic analyzer	Thermo Fisher Scientific
QIAxpert System	QIAGEN
GeneFlash Bioimaging System	Syngene
TC20 Automated cell counter	BioRad
Varioskan LUX multimode plate reader	Thermo Fisher Scientific
Trans-Blot Turbo Transfer System	BioRad
G:BOX iChemi XR5	Syngene
Leica TCS SP8	Leica Microsystems

Table 3.11 Analytical software

Analytical software	Supplier
GENESys v1.2.5.0	Syngene
LAS X	Leica Microsystems
Fiji v2.11.0	Open source
Prism GraphPad 9	Dotmatics
GIMP v2.10.34	GIMP Development Team (Open source)

4. Methods

4.1 Overview of the experimental design

This thesis is divided in two different parts according to the pathogenic CEL variant of interest. The first part covers the study of the CEL-MODY protein and the influence of cysteines on this variant's behaviour. Three different constructs based on the *CEL* gene were used (**Figure 9**): (1) CEL-WT expressing the normal *CEL* gene with 16 repeats in the VNTR region, (2) CEL-MODY-CYS encoding the *CEL* variant generated by the DEL1 mutation [68], changing the VNTR sequence and introducing 11 cysteines, one in each repeat of the tail of the protein. (3) CEL-MODY-ALA which produces a variant of the pathogenic CEL-MODY protein in which all cysteines have been substituted by alanines. All the encoded CEL proteins contain a V5 Tag (GKPIPNPLLGLDST) and a His Tag (HHHHHH). The constructs were provided by the research group of Prof. Mark Lowe, Washington University School of Medicine, St. Louis, USA. All constructs were studied by transfection into HEK293 cells and into "Simple Cells" for subsequent evaluation by western blotting of cellular extracts and immunofluorescence of fixated cells.

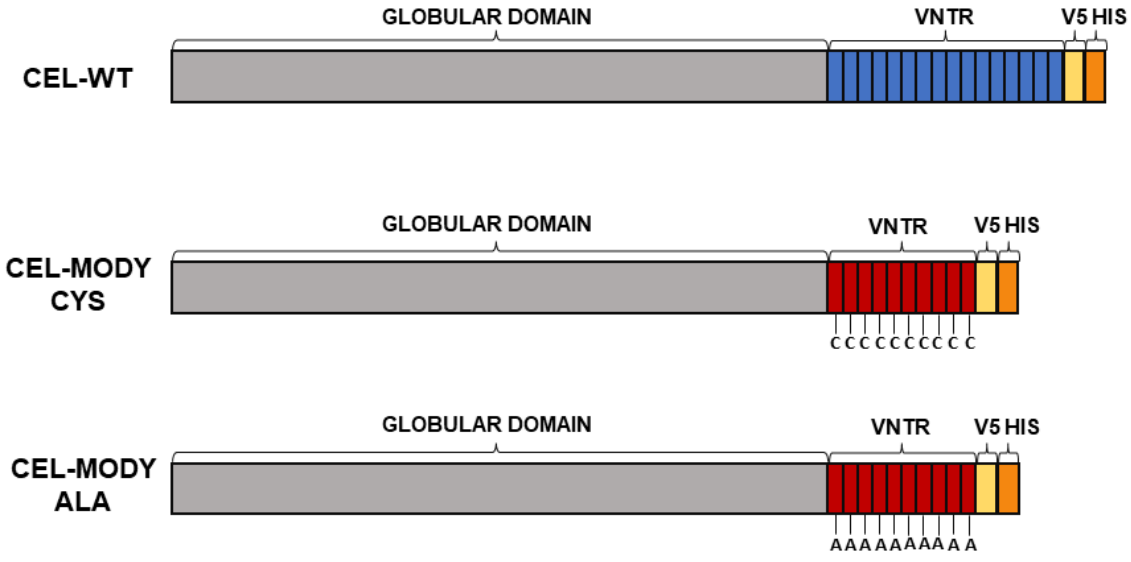

Figure 9. CEL protein variants (CEL-WT, CEL-MODY-CYS, CEL-MODY-ALA) used for assesing the effect of cysteines on the pathogenic behaviour of CEL-MODY. Grey colour shows the globular domain, yellow colour the V5 epitope tag and orange colour the HIS tag. The VNTR-encoded tails of CEL-WT and the two CEL-MODY constructs are drawn in different colours to indicate that the repeated segments of the variants differ in sequence.

The second part of the thesis is focused around the CEL-HYB1 protein. The experimental approach is similar to that of the first part. In addition to the CEL-WT construct previously explained, four different variants of *CEL-HYB1* were studied (**Figure 10**) (1) The CEL-HYB1 construct encodes the hybrid protein produced by recombined allele of *CEL* and *CELP* genes [80], which results in the introduction of two cysteines (C583, C587) in the third repeat of the CEL protein tail. (2) CEL-HYB1-C583A where C583 of CEL-HYB1 protein has been mutated to alanine. (3) CEL-HYB1-C587A where C587 has been mutated to alanine. (4) CEL-HYB-C583A/C587A where both C583 and C587 have been mutated to alanine. The three latter constructs had been generated by Khadija El Jellas using a site-directed mutagenesis approach on the CEL-HYB1 construct. All constructs were studied in the same way as the CEL-MODY constructs, except that only HEK293 cells were employed.

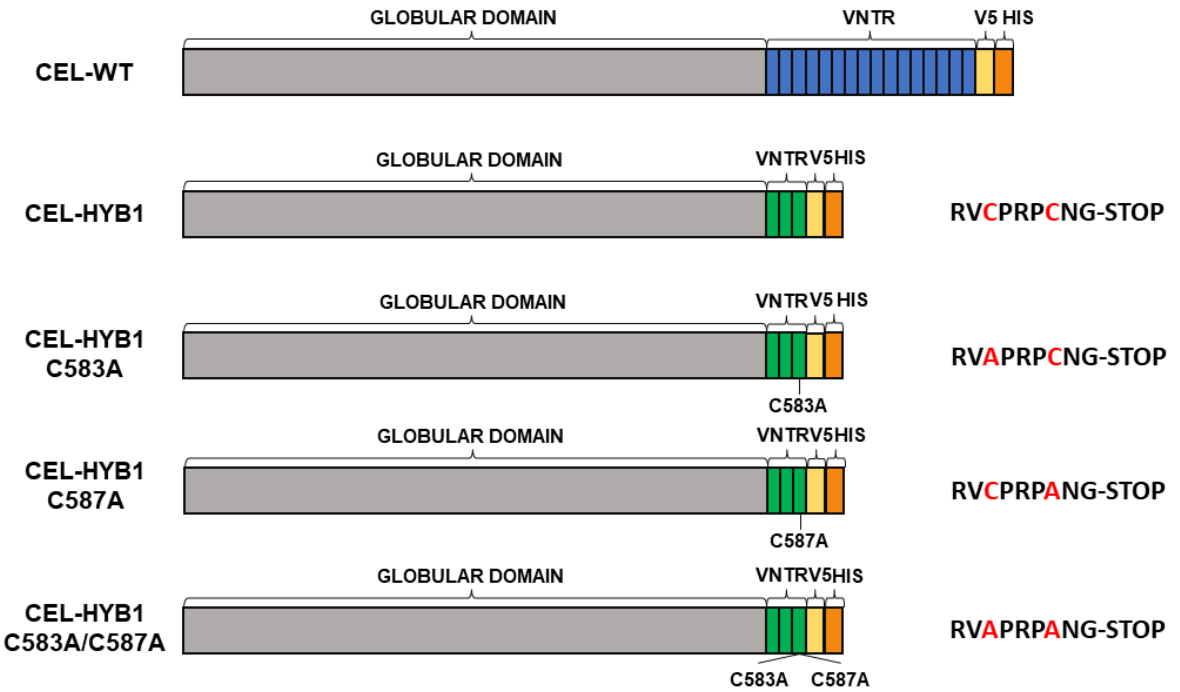

Figure 10. CEL protein variants used for assessing the effect of cysteines in the pathogenic behaviour of CEL-HYB1. The protein sequence of the last VNTR repeat (= repeat 3) of the CEL-HYB1 constructs is shown to the right. Grey colour shows the globular domain, yellow colour shows the V5 epitope tag and orange colour the HIS tag. The tail of CEL-HYB1 is drawn in green colours to indicate that the repeated segments are encoded by the CELP pseudogene.

4.2 Preparation of CEL-expressing plasmids

4.2.1 Bacterial culture and plasmid purification

E.coli transformed with each CEL variant (glycerol stocks provided by Khadija El Jellas) was plated on LB agar plates with ampicillin and incubated overnight at 37ºC. The following day a single colony was picked for each variant and used to inoculate 5 mL of LB medium containing ampicillin (100 µg/mL). This starting culture was incubated at 37 ºC with shaking (250 rpm) for 8 hours. Then 100 mL of LB medium containing ampicillin (100 µg/mL) was inoculated with 100 µL of the starting culture (1:1000 dilution). Cultures were then incubated at 37ºC with shaking (250 rpm) for 16 hours, transferred to 50 mL Falcon tubes and centrifuged at 6000 x g for 15 minutes at 4ºC. The supernatants were discarded, and bacterial pellets used for plasmid purification.

Plasmid purification was performed using QIAfilter Plasmid Maxi Kit and the protocol from the manufacturer. Buffers and consumables were provided by the kit. Bacterial pellets were resuspended in 10 mL of buffer P1 containing LyseBlue reagent. Then 10 mL of buffer P2 was added to the resuspended pellets and the mix was shaken by inversion 4 to 6 times, until the suspension had a homogenous blue colour. The suspension was incubated for 5 minutes at room temperature., followed by the addition of 10 mL chilled buffer P3 and vigorous mixing by inversion until the suspension became colourless. The suspension was poured into the barrel of a QIAfilter Cartridge and incubated at room temperature for 10 minutes. In the meantime, QIAGEN-tip 500 was equilibrated with 10 mL of buffer QBT and allowed to empty by gravity flow. Plunger was inserted into the QIAfilter Cartridge, and the cell lysate was filtered into the equilibrated QIAGEN-tips. The lysate was allowed to enter the QIAGEN-tip 500 resin by gravity flow. Once all lysate had flowed through the resin, the QIAGEN-tip was washed 2 times with 30 mL buffer QC. After washing the QIAGEN-tips, DNA was eluted into a clean 50 mL Falcon tube using 15 ml of buffer QF. DNA was then precipitated using 10.5 mL of room-temperature 70% isopropanol and centrifuged at 15000 x g for 30 minutes at 4º. The supernatant was carefully discarded. DNA pellets were then washed with 5 mL of room-temperature 70% ethanol and centrifuged at 15000 x g for 10 min. The supernatant was again carefully decanted and pellets were left to air-dry for 5-10 minutes. Finally, DNA was then redissolved in 150 µL of TE buffer and stored at 4ºC overnight. For long-term storage DNA was stored at -20ºC

4.2.2 Determination of plasmid quality and concentration

Plasmid quality was evaluated via agarose gel electrophoresis. 1% agarose gel was used. 200 ng of plasmid were loaded in each well. Concentration of purified plasmids was measured using the QIAxpert System. Absorbance of 2 µL of sample was measured at 260 nm. A 260/280 ratio of around 1.8 and 260/230 ratio of between 2.0 and 2.2 were accepted as signs of pure sample. DNA was stored at -20 ºC until further use.

4.2.3 Agarose gel electrophoresis

Purified plasmid quality was further verified via agarose gel electrophoresis. Samples were separated on a 0.8% agarose gel containing 0.625 µg/µl ethidium bromide. 5 µl of plasmid (200 ng DNA) was mixed with 3 µl of 6x loading buffer before loading on the gel. 1 kb DNA ladder was added as a molecular-weight marker. The gel was run for approximately 1 hour and 20 minutes at 80V in 1x TBE buffer. Fluorescence visualization was performed using the Gene Flash Bioimaging system.

4.2.4 Sanger Sequencing

Plasmids were sequenced to verify that they contained the desired modifications in the VNTR regions. For this, Sanger sequencing was used. Chain-termination PCR mastermix, primers and program are described in table 4.1, 4.2 and 4.3 respectively. Chain-termination PCR was performed on Applied Biosystems 2720 Thermal Cycler. The chain-termination PCR product was then cleaned using Sephadex G-50 that had been prepared the day before. For this, an MS-HV plate was loaded with Sephadex using a Multiscreen 45 µL Column Loader. 300 µL of ddH_2O was added to each well. The MS-HV plate was stored overnight at 4 ºC. The next day the MS-HV plate was centrifuged at 910 x g for 5 minutes and chain-termination PCR products were added. The plate was centrifuged at 950 x g for 5 minutes. Samples were collected in a 96-well microtiter plate and sequenced on a 3500xL Genetic Analyzer.

Table 4.1 Contents of chain-termination PCR mastermix

Component	Volume (µL)
BigDye v.3.1	1.0
5X Big Dye Buffer	1.0
Betaine	2.0
Primer (20 µM)	0.25
Plasmid	1.0
ddH_2O	4.75

Table 4.2 Primers used for Sanger sequencing of plasmids

Primer	Binding site in *CEL* gene	Sequence (5' → 3')
T7 (forward)	T7 promotor	ATTATGCTGAGTGATATCCC
DR (reverse)	Exon 11	GCCGCTGTTTTCCGTA
EF (forward)	Exon 11	CACACACTGGGAACCCT
BGH (Reverse)	BGH polyadenylated sequence	ATCTTCCGTGTCAGCTCC

Table 4.3 PCR thermal cycling program for Sanger sequencing

	Temperature (ºC)	Time	Number of cycles
Pre-heat	96	1 min	
Denaturation Annealing Elongation	96 58 60	10 sec 5 sec 4 min	25
Hold	4	∞	

4.3 Cell culture and transfection

HEK293 cells [134] and "Simple Cells" (SC) [133] (HEK 293 cells with "simplified" O-glycosylation machinery) were used for cellular experiments described in this thesis.

4.3.1 Thawing

HEK293 and SC were thawed quickly in a 37 ºC water bath. 1 mL of prewarmed complete DMEM was added to acclimate cells and they were then transferred into a 15 mL tube containing 8 mL of pre-warmed Dulbecco's modified Eagle's medium (DMEM) + GlutaMAX^{TM} supplemented with 10% FBS and 100 U/mL Antibiotic-Antimycotic solution (complete DMEM). Cells were centrifuged at 300 x g for 5 minutes. The supernatant was carefully removed. Cells were resuspended in 1 mL of complete pre-warmed DMEM and transferred into a T75 flask containing 12.5 mL of complete pre-warmed DMEM.

4.3.2 Sub-culturing and seeding

For passage of HEK293 and SC, DMEM was removed, and cells were washed in pre-warmed Dulbecco's phosphate buffered saline (DPBS). DPBS was removed and 3 mL of pre-warmed 0.05% Trypsin-EDTA was added to the cells. Cells were incubated at 37 ºC for 3 minutes. After incubation, 7 mL of pre-warmed complete DMEM was added to stop trypsinization and cells were transferred to a 15 mL tube. Cells were centrifuged at 500 x g for 5 minutes. The

supernatant was discarded, and the cells were resuspended in 1 mL of complete DMEM. 100-200 µL of the resuspended cells was transferred to a T75 flask with 12.5 mL of prewarmed complete DMEM. Cells were then incubated at 37 ºC in a 5% CO_2 atmosphere until next passage.

4.3.3 Freezing

Cells were grown until 70-80 % confluency in T75 flasks. DMEM was removed and cells were washed in pre-warmed DPBS. DPBS was removed and 3 mL of pre-warmed 0.05% Trypsin-EDTA was added to the cells. Cells were incubated at 37 ºC for 3 minutes. After incubation, 7 mL of pre-warmed complete DMEM was added to stop trypsinization and cells were transferred to a 15 mL tube. Cells were centrifuged at 500 x g for 5 minutes. The supernatant was removed, and cells were then resuspended in 3 mL of freezing medium (10 % DMSO diluted in DMEM + 50 % FBS). 1 mL of resuspended cells was then transferred into each freezing vial. The vials were transferred to a CoolCell™ LX Freezing Container and stored at -80 ºC for one day. For long time storage cells were placed in a liquid nitrogen tank.

4.3.4 Transient transfection

Prior to transfection, cells were seeded in 6-wells plates ($5x10^5$ cells per well) or in 12-well plates ($1.5x10^5$ cells per well) and grown for 24 hours in DMEM + GlutaMAX™ supplemented with 10% FBS and no Antibiotic-Antimycotic solution.

HEK293 and "Simple cells" were transfected using Lipofectamine® 3000 following the manufacturer's instructions. 1 and 2.5 ug of DNA were used for the transfection of cells seeded in 12-well and 6-well plates respectively and 2 µL of lipofectamine was used per 1 µg of DNA (2 µL for 12-well plates and 5 µL for 6-well plates). Lipofectamine® 3000 was diluted in OPTIMEM in one tube while DNA and P300 reagent were diluted in OPTIMEM in a different tube. The mixture of DNA and P300 reagent was then added into the tube containing lipofectamine. The mix was incubated for 10-15 minutes at room temperature. The DNA-lipid mix was then added to the cells, which were incubated at 37 ºC for 4 hours before cellular media was changed. The cells were further grown for 48 hours at 37 ºC and 5 % CO_2 until collection for western blot or fixation for immunocytochemistry.

4.4 Cell fractionation

Prior to preparation of samples for western blotting, HEK293 cells and SC cells were seeded in 6-well plates (5 x 10^5 cells per well) and grown overnight. Cells were then transfected using Lipofectamine 3000® as described previously (**Section 4.3.4**) and after 48 hours cells were lysed and separated into different fractions. Three independent transfections were performed.

4.4.1 Preparation of cell fractions (medium and cell lysate/pellet fractions)

48 hours post transfection, 1 ml of cell growth medium was collected and centrifuged at 14000 x g in order to remove any detached cells. 750 µL of the supernatant was collected and used for further analysis, referred to as the 'medium'.

For detection of intracellular proteins, cells were washed with prewarmed PBS and 150 µL of ice-cold RIPA lysis buffer containing protease inhibitors was added to each well. Using a cell scraper, cells were detached from the plate surface. RIPA buffer containing scraped cells was transferred into 1.5 mL tubes and incubated on ice for 30 minutes. Next, the tubes were centrifuged at 17000 x g for 15 minutes at 4ºC. The supernatant containing the soluble proteins of the cell lysate was carefully removed and transferred into clean 1.5 mL tubes. These samples will be referred to as the 'cell lysate' fraction.

The remaining pellet, containing the detergent-insoluble proteins, was washed 2 times with 200 µL of ice-cold PBS and centrifuged at 17000 x g for 15 minutes at 4ºC. The supernatant was discarded and 100 µL of 2x LDS loading buffer was added to the dry pellet. These samples will be referred to as the 'pellet' fraction. All fractions were stored at -80 ºC until further use.

4.4.2 Determination of protein concentration

The protein concentration of the cell lysate fraction was measured using Pierce™ BCA Protein Assay kit according to the instructions from the manufacturer. In brief, 10 µL of sample were loaded in a 96 well plate. Then 200 µl of BCA working reagent was added to each sample (50 parts reagent A: 1 part reagent B). Serial dilutions of albumin ranging from 62.5 µg/mL to 2000 µg/mL were used to construct a standard curve. The plate was then incubated at 37ºC for 30 minutes. For detection, Varioskan™ LUX multimode microplate reader was used in conjunction with SkanIt™ Software to measure the absorbance at 562 nm.

4.5 SDS-PAGE and Western blotting

4.5.1 Sample preparation

Samples from the medium and the cell lysate were prepared for separation on the polyacrylamide gels as described in Table 4.4. Volume of sample needed was calculated using the protein concentration measured for the lysate fraction in Section 4.4.2. Same volume was used for medium and lysate fractions. Samples where then denatured at 95ºC for 5 minutes.

Table 4.4 Sample preparation of medium and cell lysate fraction

Component	Volume/Quantity
10x Reducing Agent (DTT)	4 µL
Sample	20 µg
4x LDS sample buffer	4 µL

Samples from the pellet needed a more elaborate preparation method. The pellet diluted in 100 µL of 2x LDS loading buffer was incubated at 70 ºC for 20 minutes, vortexing the sample for 10 seconds every 4 minutes to separate insoluble proteins from cellular membranes. After incubation, samples were centrifuged at 17000 x g for 1 minute and the same sample volume used for the cell lysate and medium was used, following the volumes and quantities described in Table 4.5. Samples where then denatured at 95ºC for 5 minutes.

Table 4.5 Sample preparation of pellet fraction

Component	Volume/Quantity
Sample	Same volume as cell lysate and medium fractions
10x Reducing Agent	4 µL

4.5.2 SDS-PAGE

Samples were then loaded on a NuPAGE 10% Bis-Tris Gel using XCell SureLock Mini-Cell Electrophoresis System. 4 µL of Precision Plus Protein Dual Color Standards were used as protein size marker. 1x MOPS buffer was used as electrophoresis buffer and 500 µL of NuPAGE Antioxidant was added to each chamber. 100V was applied for 15-20 minutes, and then increased to 120V until the protein marker for 25 kDa had reached the bottom of the gel.

4.5.3 Western blotting

Following SDS-PAGE the proteins were transferred from the gel to a polyvinylidene fluoride (PVDF) membrane using the trans-Blot Turbo Transfer System. For this, gels were removed from their plastic cases and rinsed with dH_2O. The bottom part of the Trans-Blot Turbo Mini/Midi PVDF Transfer pack was placed on the transferring cassette and rolled down to eliminate bubbles. The gel was then added on top of the membrane and lightly rolled to eliminate bubbles. The top part of the transfer pack was added and rolled down. The cassette was closed and introduced into the machine. Midi transfer packs blotting was performed at 2.5 A constant current and up to 25 V voltage for 8.5 minutes. Mini transfer packs were blotted at 1.3 A and up to 25 V for 8.5 minutes. After blotting, the membrane was blocked in 5% BSA diluted in PBS with 0.1% Tween20 for 1 hour on a tilting shaker. The blocking solution was then discarded, and primary antibody diluted in 3% BSA was added. The membrane was incubated overnight at 4ºC on a tilting shaker. Next day, the primary antibody was removed, and the membrane washed 5 times in PBS-T 0.1% for 5 minutes. Secondary antibody diluted in 3% BSA was then added, and the membrane was incubated for 1 hour at room temperature on a tilting shaker. The antibody solution was removed, and the membrane was washed as for the primary antibody. Protein bands were visualised using SuperSignal™ West Femto Maximum Sensitivity Substrate. For detection, Syngene G:BOX iChemi XR5 imager and GENESys software were used.

4.5.4 Relative quantification of western blotting

For relative protein amount quantification, the digital image data from the membrane was analysed using the open-source image software Fiji [135]. The results were normalized to the intensity of the protein band representing GAPDH (used as total protein control). Each western blot experiment was repeated three times.

4.6 Immunocytochemistry and confocal microscopy

HEK 293 cells were seeded in 12-well plates (1.5×10^5 cells per well) containing glass coverslips and grown overnight. Prior to cell seeding, the coverslips were coated with 1% ECM and 2 μg/mL fibronectin diluted in DMEM containing 100 IU/mL penicillin and 100 μg/mL streptomycin for 1 hour to increase cell adherence on glass. The coating solution was then removed, and cells were seeded. Once seeded, cells were left to grow overnight. Cells were then transfected using Lipofectamine 3000® as described in **Section 4.3.4** and cultured for 48 h.

4.6.1 Fixation and immunostaining

Two days after transfection, the cells were rinsed with pre-warmed DPBS. They were fixed for 30 minutes in 500 µL 4% paraformaldehyde, 0.2 M phosphate buffer, pH 7.2. Fixation was stopped by adding 500 µL of 1 M glycine. Cells were then washed with PBS and incubated for 15 minutes in 500 µL 100 mM glycine. Afterwards, glycine was removed by aspiration and 500 µL of 50 mM NH_4Cl was added to the cells for 15 minutes to block unreacted aldehyde groups. After removal of NH_4Cl, cells were permeabilized using 500 µL of 0.1% Triton-X100 diluted in PBS for 15 minutes. Cells were cleaned with PBS twice. 500 µL of 5% goat serum diluted in PBS was added as blocking solution, and the blocking was continued for 30 minutes. Cells were stained with primary antibody overnight at 4º C in a humidity chamber. Primary antibody was diluted in blocking solution. Next day, cells were washed 2 x 10 minutes in washing solution (PBS + 0.1% Tween20). A secondary antibody was then added and incubated for 1 hour at room temperature in a humidity chamber. The secondary antibody was also diluted in blocking solution. After incubation cells were washed 2 times for 10 minutes with washing solution. Then Hoechst diluted in PBS (1 µl in 10 mL) was added and incubated for 5 minutes. Cells were rinsed with PBS and with ddH_2O before mounting. For mounting, coverslips were lightly dried and inverted onto glass slides in a drop of ProLong Glass Antifade mounting reagent. Slides were left at room temperature for 24 hours in the dark and then stored at -20ºC.

4.6.2 Confocal imaging

Finished slides were examined with a Leica TCS SP8 MP confocal microscope (Leica Microsystems) with a 100x HC PL APO, 1.40NA, STED WHITE oil objective. Images were acquired with Diode 405 and WLL lasers using LAS X Life Science Software. The imaging was performed at the Molecular Imaging Center (MIC), Department of Biomedicine, University of Bergen.

Images were processed and analysed with Fiji software and BIOP JACoP plugin to obtain the colocalization values [136]. For this, a region of interest was created around every cell of interest to individualize the analysis. Then BIOP JACoP was used to obtain the Pearson's coefficients of each stack, measuring the correlation between green and red pixels. Stacks of four images per cell were selected (0.3 µm between images), trying to always select the same regions for each cell (Including nucleus, strong intensity of CEL). The Pearson's coefficient for each image was used to determine the coefficient of each cell as a mean of the four images

on each stack. Processed images were structured in the figures using Gimp image software [137].

4.7 Statistical analysis

All statistical analysis were performed using GraphPad Prism 9 [138]. The significance of differences was tested by one-way ANOVA. Multiple comparisons between individual groups were performed via Turkey's multiple comparisons test. $P \leq 0.05$ was considered statistically significant (* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$).

5. Results

5.1 Plasmids preparation and sequencing

Prior to the analysis of the different CEL variants in cellular systems, bacterial cultures containing the plasmids to be used in the study (CEL-WT, CEL-MODY-CYS, CEL-MODY-ALA, CEL-HYB1, CEL-HYB1-C583A, CEL-HYB1-C587A, CEL-HYB1-C583A/C587A) were grown as described in **Section 4.2.1.** Plasmid DNA was then isolated using the QIAfilter Maxi kit (**Section 4.2.1**). Spectrophotometry was used both for quality control ($A_{260/280}$ ratio) and to measure plasmid concentration (**Section 4.2.2**). Plasmid quality was further evaluated via agarose gel electrophoresis (**Figure 11**). The lowermost DNA bands correspond to the supercoiled form of the plasmids. All plasmids migrated with similar molecular mass and had comparable band intensities, showing that the quality of the plasmids preparations was satisfactory. To verify the sequence of the VNTR region for each construct, Sanger sequencing was performed. Results are shown in **Appendix A.**

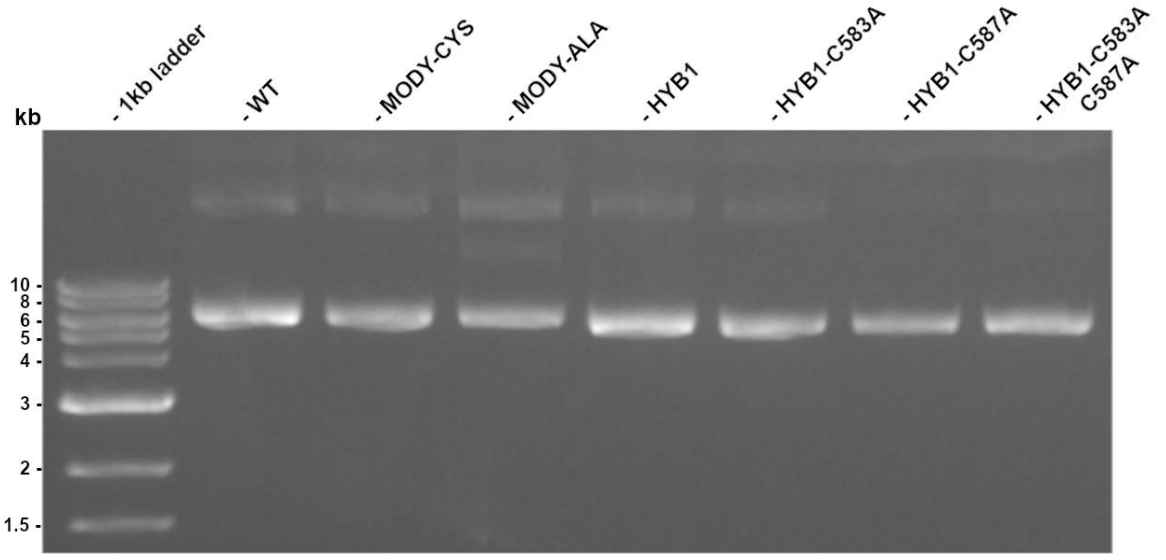

Figure 11. Agarose gel electrophoresis of CEL plasmids. The plasmids contain the different CEL constructs used in the thesis. 200 ng of DNA were loaded on a 1% agarose gel prepared in 1X TBE buffer and EtBr

5.2 Assessment of Cys residues in the CEL-MODY variant

5.2.1 Secretion and protein distribution

To assess the effect of cysteine (Cys) residues on protein secretion and intracellular localization of the CEL-MODY protein, HEK293 cells were transiently transfected with plasmids expressing CEL-WT, CEL-MODY-CYS and CEL-MODY-ALA. Forty-eight hours after transfection, cell medium was collected, and the lysate and pellet fractions were isolated. The samples were analysed via SDS-PAGE and western blotting (**Figure 12A**). Immunodetection of CEL was performed using a monoclonal antibody against the C-terminal V5 tag (anti-V5 antibody). GAPDH staining was performed to monitor the loading of the gel and was also used for normalisation. CEL-WT was detected as a band between 100 and 120 kDa and CEL-MODY-CYS and CEL-MODY-ALA around 75 kDa. When comparing CEL-MODY-CYS and CEL-MODY-ALA in the medium fraction, the CEL-MODY-ALA band was seen at a slightly higher molecular weight. Moreover, the blots from the lysate fraction showed different patterns in how the signal was detected, CEL-WT and CEL-MODY-ALA appeared as two clear bands separated by approximately 20 kDa while CEL-MODY-CYS presented as one band and a low-intensity smear. Finally, for the pellet fraction the CEL-MODY-CYS band was present as a relatively strong smear, while CEL-MODY-ALA and CEL-WT showed as two bands with almost identical molecular weight.

Three independent experiments were performed. Quantification of the immunoblots is shown in **Figure 12B**. The graph for the medium fraction shows that CEL-MODY-ALA had higher signal intensity than CEL-WT, with CEL-MODY-CYS exhibiting the weakest signal. For the lysate fraction, CEL-WT and CEL-MODY-ALA had similar signal intensities while CEL-MODY-CYS appeared as a weaker band. Finally, for the pellet fraction, CEL-MODY-CYS was the most abundant variant, exhibiting a signal that was more than double the intensity of the CEL-MODY-ALA signal. CEL-WT was present only as a very weak band. For the medium and pellet fractions all variants displayed statistically significant differences, while for the lysate fractions only trends could be observed.

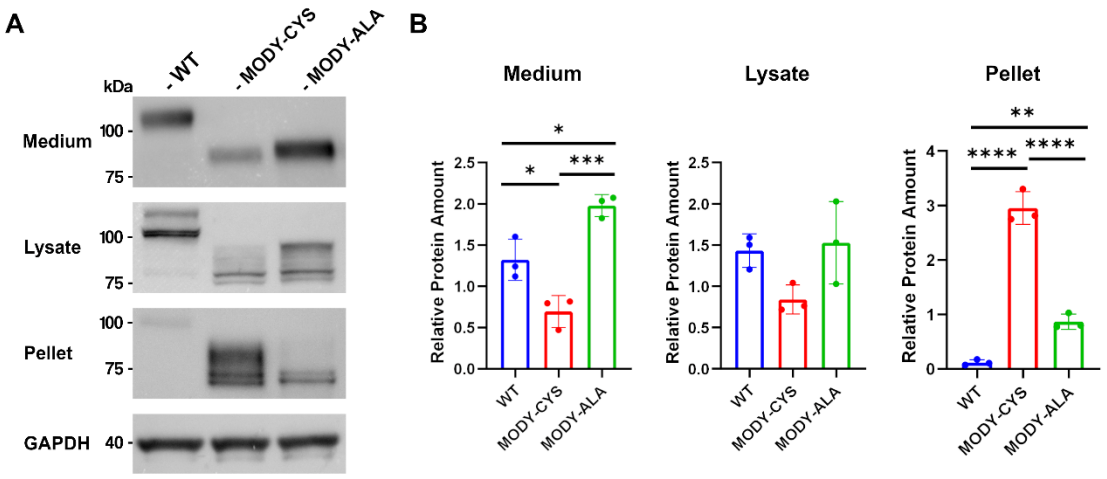

Figure 12. Secretion and intracellular protein distribution of CEL variants in transfected HEK293 cells. (A) Cells were transfected with the CEL-WT, CEL-MODY-CYS and CEL-MODY-ALA constructs. SDS-PAGE and western blot were used to analyse the medium, lysate and pellet fractions using Anti-V5 antibody. Blots shown are representative of three independent experiments. GAPDH was used as loading control (B) Quantification of western blot signal intensities of the three experiments normalized according to the GAPDH signal. Error bars represent standard deviation (SD). Statistical significance was assessed using one-way ANOVA analysis where * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$.

5.2.2 Endoplasmic reticulum stress markers analysis via western blot

Cysteine residues can lead to aggregation and misfolding (**Section 1.6**). To evaluate if the cysteines of CEL-MODY could induce changes in the biochemical properties that lead to ER stress, the pellet fractions isolated in **Section 5.2.1** were again analysed via SDS-PAGE and western blotting (**Figure 13A**). Immunodetection was now done for calnexin (an ER chaperone involved in the folding of N-glycosylated proteins such as CEL [139]); BiP (an ER chaperone central in the unfolded protein response [140]) and PERK (an ER stress signal transducer that directly interacts with BiP [140]). Specific antibodies against each protein in conjunction with HRP-conjugated secondary antibody were employed. GAPDH was analysed in the lysate fraction and included as a loading control. All markers were located at their expected molecular weights of 75 kDa for calnexin and BiP, and 140 kDa for PERK. PERK was present at very low signal intensity and therefore was hard to quantify.

Three independent replicates of each ER stress marker were performed. Band intensity was quantified, and results were graphed (**Figure 13B**). For all three markers no statistically significant differences were seen between the different variants. The strongest trend was for the BiP marker, where CEL-WT presents slightly lower levels than CEL-MODY-CYS and CEL-MODY-ALA.

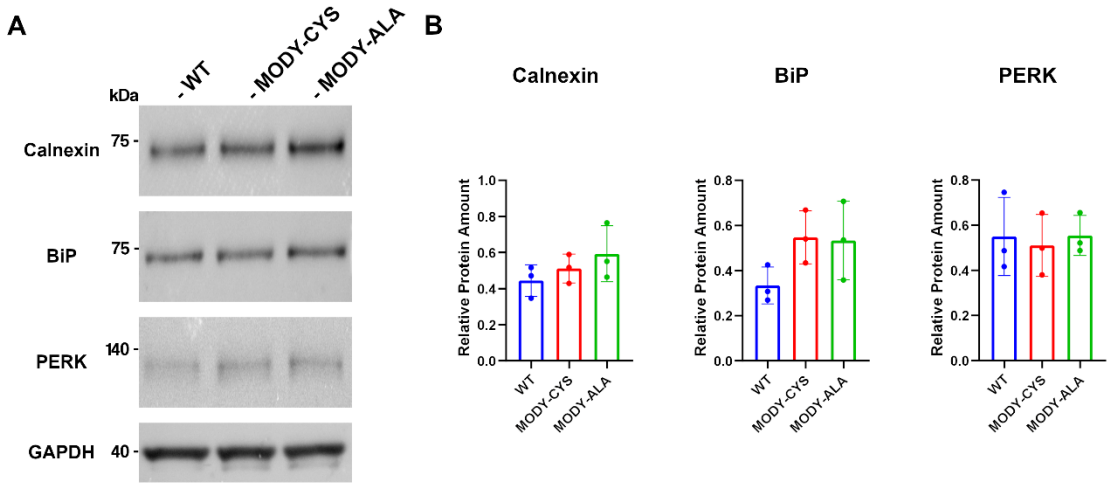

Figure 13. Protein abundance of endoplasmic reticulum stress markers in transfected HEK293 cells. (A) Cells were transfected with the CEL-WT, CEL-MODY-CYS and CEL-MODY-ALA constructs. SDS-PAGE and western blot were used to analyse pellet fractions. Blots shown are representative of three independent experiments. GAPDH was measured in the lysate fraction and used as loading control (B) Quantification of chemiluminiscence intensity of the ER stress markers; three independent experiments were performed and GAPDH signal was used for normalizing. Error bars represent standard deviation (SD). Statistical significance was assessed using one-way ANOVA analysis where * $P \leq 0.05$.

5.2.3 CEL intracellular protein localization via immunocytochemistry

HEK293 cells transiently transfected with CEL-WT, CEL-MODY-CYS and CEL-MODY-ALA proteins were fixed and immunostained to investigate intracellular CEL distribution and its relationship with different ER stress and Golgi markers. CEL was visualised using the anti-V5 antibody and cells were then co-stained with one of the ER stress markers calnexin and BiP, or the Golgi marker GALNT2, an N-acetylgalactosaminyltransferase involved in O-glycosylation within the Golgi apparatus [141]. The nuclei were visualized by using DAPI.

Confocal images of the CEL protein variants co-stained with calnexin, BiP or GALNT2 are shown in **Figure 14**, **Figure 15** and **Figure 16**, respectively. Green signal represents the intracellular distribution of CEL whereas the red signals mark the other proteins. For CEL-WT and CEL-MODY-ALA the intracellular distribution was characterized by localized regions resembling the distribution of the Golgi apparatus. However, CEL-MODY-CYS displayed a different distribution, showing a more network-like pattern, similar to the ER. In addition, in the CEL-MODY-CYS slides there was extracellular dot-like staining possibly representing extracellular protein aggregates.

Regarding the ER stress markers, both calnexin and BiP (**Figure 14** and **Figure 15**) presented no differences regarding the distribution of these markers within the cell. However, CEL MODY-CYS showed a higher degree of colocalization with these two markers compared to CEL-WT and CEL-MODY-ALA, represented by more yellow colour being present in the merged images. For GALNT2 (**Figure 16**), no differences were observed regarding its intracellular localization. CEL-WT and CEL-MODY-ALA displayed a larger yellow area in the merged images, indicating that more protein is present in the Golgi compartment for these variants.

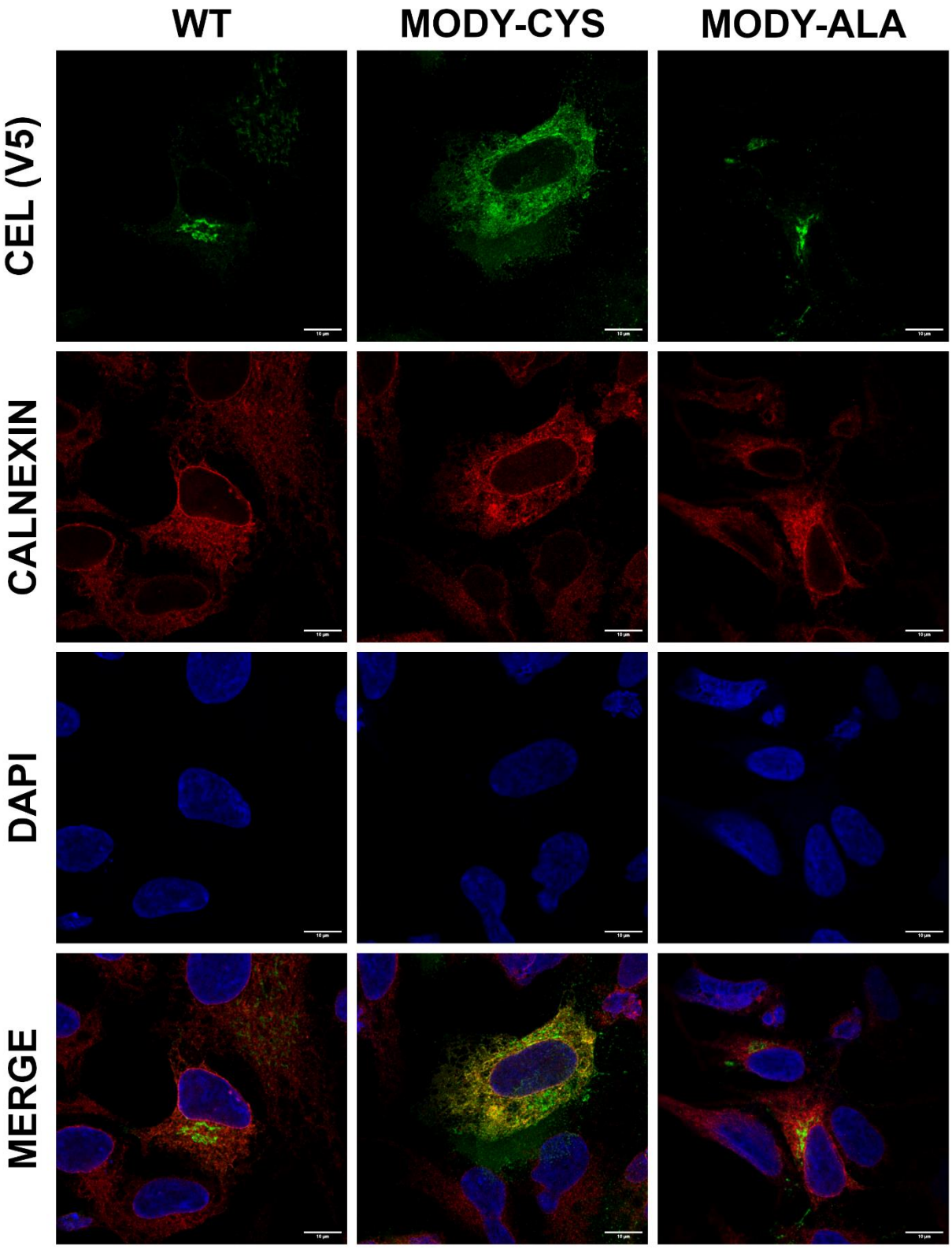

Figure 14. Intracellular distribution of CEL and calnexin in HEK293 cells. Cells were transfected with CEL-WT, CEL-MODY-CYS and CEL-MODY-ALA, fixed and stained for CEL (Anti-V5, green fluorescence) and Calnexin (red fluorescence). Nuclei are stained blue with DAPI. Images were acquired in a confocal microscope at 100X magnification. Scale bar = 10µm.

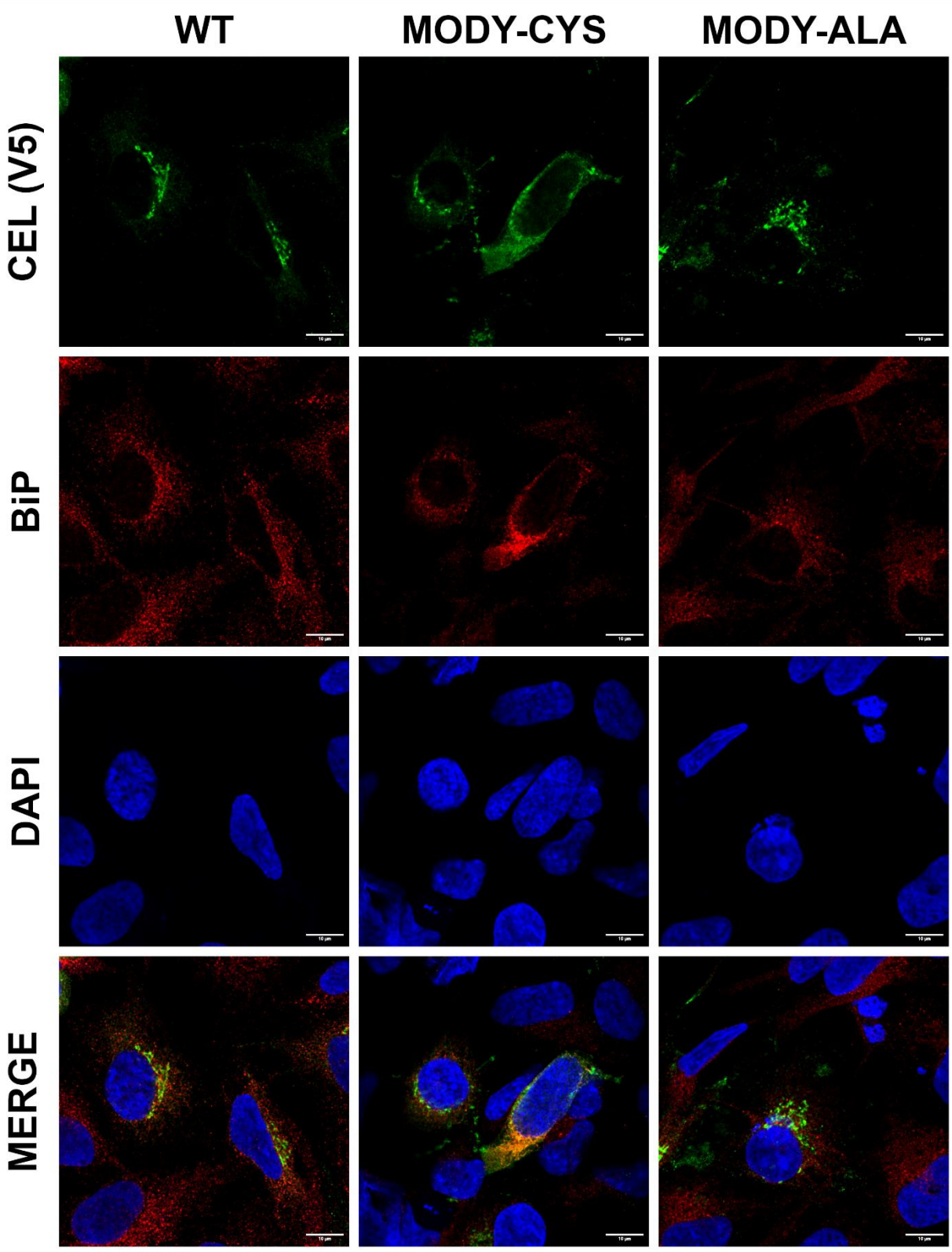

Figure 15. Intracellular distribution of CEL and BiP in HEK293 cells. Cells were transfected with CEL-WT, CEL-MODY-CYS and CEL-MODY-ALA, fixed and stained for CEL (Anti-V5, green fluorescence) and BiP (red fluorescence). Nuclei are stained blue with DAPI. Images were acquired in a confocal microscope at 100X magnification. Scale bar = 10µm

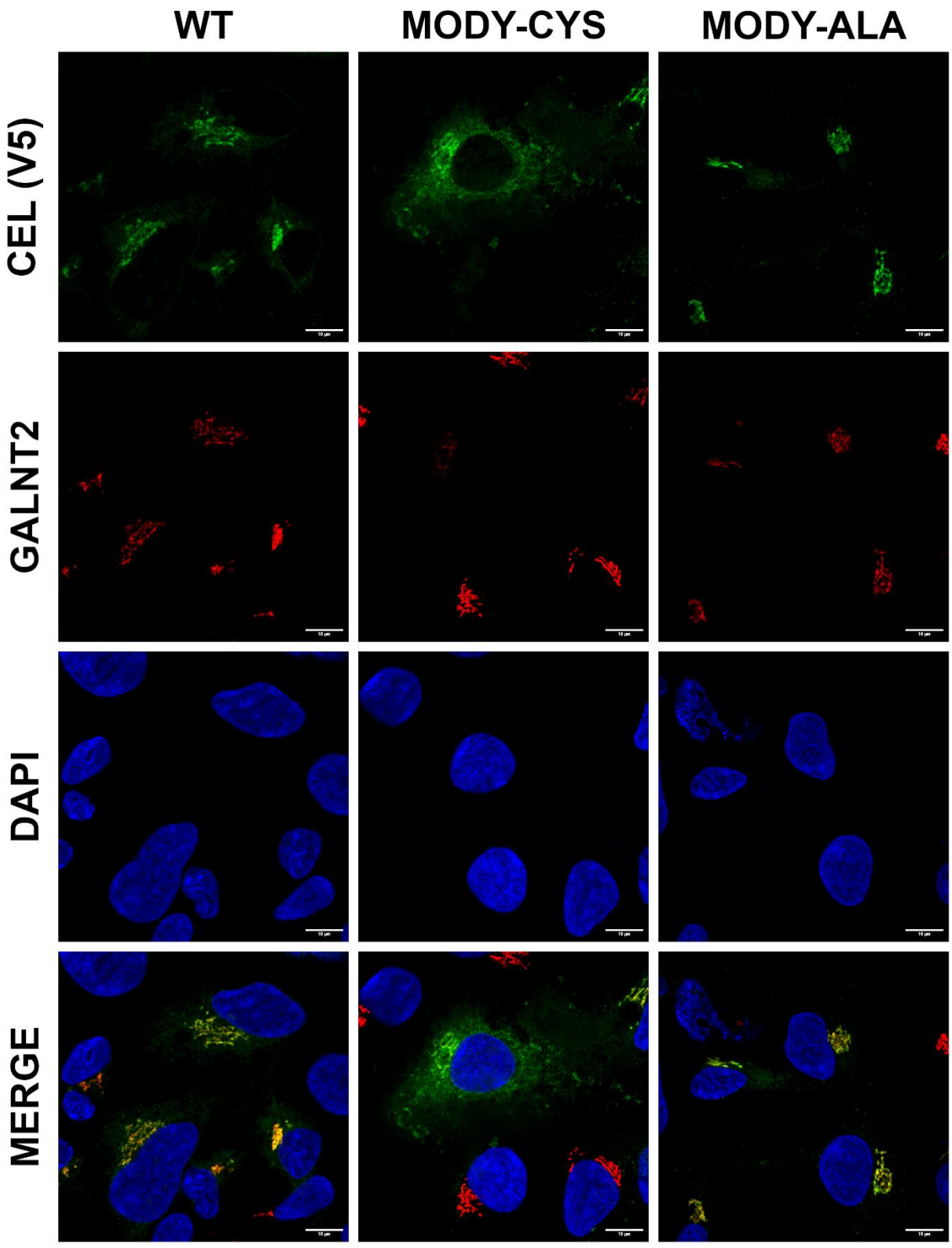

Figure 16. Intracellular distribution of CEL and GALNT2 in HEK293 cells. Cells were transfected with CEL-WT, CEL-MODY-CYS and CEL-MODY-ALA, fixed and stained for CEL (Anti-V5, green fluorescence) and GALNT2 (red fluorescence). Nuclei are stained blue with DAPI. Images were acquired in a confocal microscope at 100X magnification. Scale bar = 10µm

To automatically assess the colocalization of CEL with calnexin, BiP and GALNT2, the Pearson's coefficient was calculated using JaCoP plugin in Fiji, with higher Pearson's coefficient representing a higher degree of colocalization. The coefficients were calculated for five different cells, and a stack of four images from each cell was used to calculate the mean value. The data from this analysis is presented in **Figure 17**. The calculations showed that CEL-MODY-CYS colocalized more with Calnexin and BiP, therefore being more abundant in the ER than the two other variants. CEL-WT and CEL-MODY-ALA presented a higher degree of colocalization with GALNT2, hence they are more present inside the Golgi than CEL-MODY-CYS. Differences were statistically significant when comparing CEL-MODY-CYS to both CEL-WT and CEL-MODY-ALA. Differences between CEL-WT and CEL-MODY-ALA did not reach statistical significance.

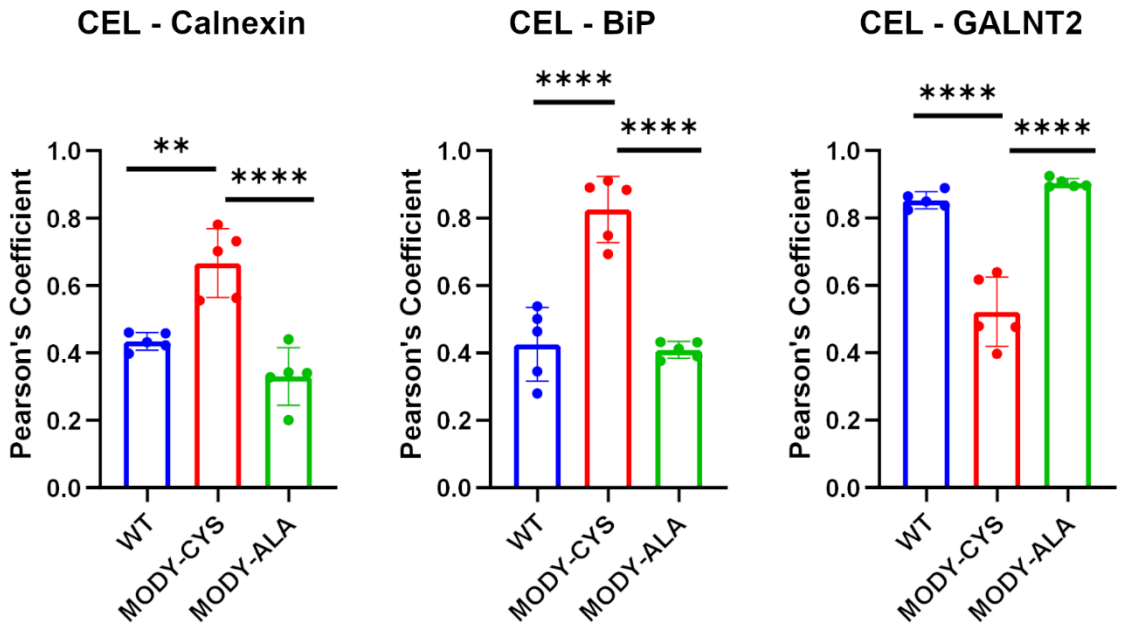

Figure 17. Colocalization of CEL protein variants with ER stress and Golgi markers. Pearson's coefficients were calculated based on immunocytochemistry images from HEK293 cells co-stained for CEL and Calnexin, BiP or GALNT2. Pearson's coefficients were calculated for 5 cells per variant and 4 stacks of each cell were used to calculate the mean. Error bars represent standard deviation (SD). Statistical significance was assessed using one-way ANOVA analysis where ** $P \leq 0.01$, **** $P \leq 0.0001$.

5.2.4 Analysis of O-glycosylation

In order to investigate whether the presence of cysteine residues influences O-glycosylation of the CEL-MODY protein, HEK293 cells and "Simple Cells" (SC = HEK293 cells with a disrupted gene encoding the COSMC chaperone) were transiently transfected with the CEL-WT, CEL-MODY-CYS and CEL-MODY-ALA constructs. Forty-eight hours after transfection, growth medium was collected, and the lysate and pellet fractions were isolated. The samples were analysed via SDS-PAGE and western blotting (**Figure 18A**). Immunodetection of CEL was performed with the Anti-V5 antibody. GAPDH staining was performed to ensure equal protein loading. For HEK293 cells, all variants migrated with the same molecular weight as in **Section 5.2.1**. However, when looking at the SC, all variants in the medium fraction showed ~20 kDa lower molecular weight, probably as result of having a version of each protein with drastically reduced O-glycosylation. In addition, when observing the lysate fraction, CEL-WT and CEL-MODY-ALA presented a different band pattern when expressed in SC. Furthermore, all variants showed higher signal intensity when transfected into SC.

Three independent experiments were performed. Signal intensities of the western blot protein bands were quantified, and the results are shown in **Figure 18B**. When comparing the cellular distribution of CEL in HEK293 and SC, despite the differences in intensity, all variants maintained a similar distribution in both the medium and lysate fractions, similar to the distribution described in **Section 5.2.1**. CEL-WT and CEL-MODY-ALA had a similar abundance and CEL-MODY-CYS showed approximately 50% reduction of signal intensity compared to the two other variants, although these differences were only statistically significant in SC as the signal from CEL-MODY-CYS is slightly weaker. In the pellet fraction the distribution changed between the two cellular types and SC showed an increased abundance of both CEL-WT and CEL-MODY-ALA. In this case, the CEL-WT signal increased to around half the signal of CEL-MODY-CYS whereas the CEL-MODY-ALA signal appeared almost identical to that of CEL-MODY-CYS.

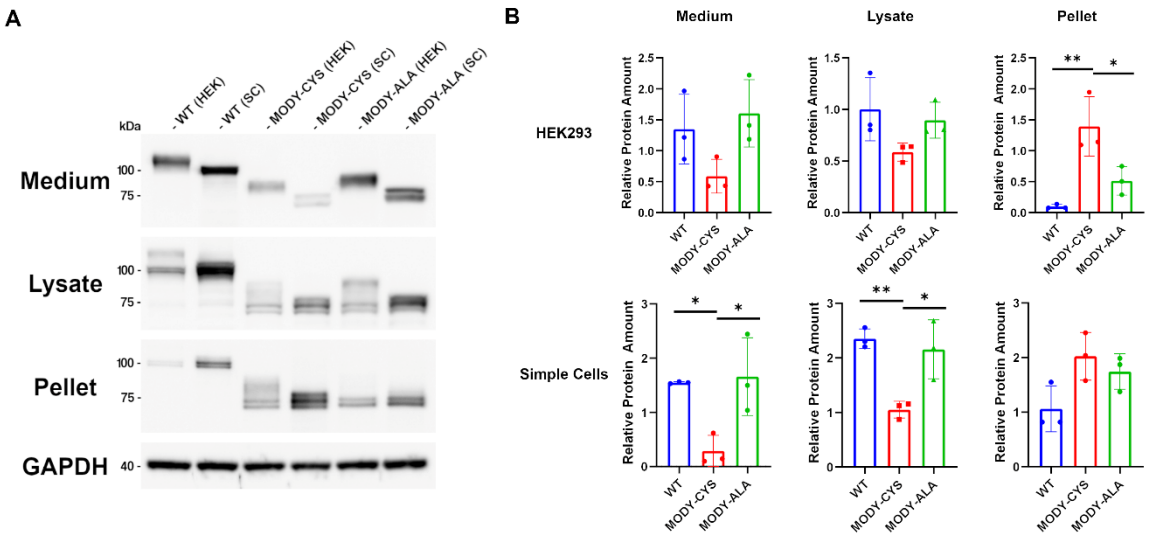

Figure 18. Secretion and intracellular protein distribution of CEL variants in transfected HEK293 cells and «Simple cells» (SC). (A) Cells were transfected with the CEL-WT, CEL-MODY-CYS and CEL-MODY-ALA constructs. SDS-PAGE and western blot were used to analyse the medium, lysate and pellet fractions using Anti-V5 antibody. Blots shown are representative of three independent experiments. GAPDH was used as loading control (B) Quantification of western blot signal intensities of the three experiments normalized according to GAPDH signal. Error bars represent standard deviation (SD). Statistical significance was assessed using one-way ANOVA analysis where * $P \leq 0.05$, ** $P \leq 0.01$.

5.3 Assessment of Cys residues in the CEL-HYB1 variant

5.3.1 Secretion and protein distribution

HEK293 cells were transiently transfected with plasmids expressing CEL-WT, CEL-HYB1, CEL-HYB1-C583A, CEL-HYB1-C587A and CEL-HYB1-C583A/C587A to assess the effect of cysteine residues on protein secretion and intracellular localization of the CEL-HYB1 protein. Forty-eight hours after transfection, cell medium was collected, and the lysate and pellet fractions were isolated as described earlier. The samples were analysed via SDS-PAGE and western blotting (**Figure 19A**). Immunodetection of CEL was performed using the anti-V5 antibody. GAPDH staining was performed to monitor the loading of the gel. CEL-WT was again detected as a band between 100 and 120 kDa and CEL-HYB1 variants were detected between 50 and 60 kDa. In the medium fraction, CEL-WT was present as single band while all CEL-HYB1 variants manifested as a smear of low signal intensity. Moreover, the blots from the lysate fraction showed different patterns in how the signal was detected, CEL-WT appeared as two clear bands separated by approximately 20 kDa, as observed in **Section 5.2.1**,

while all CEL-HYB1 variants were visualized as two closely separated bands. Finally, for the pellet fraction, CEL-WT only showed two very close bands of low signal intensity, while all CEL-HYB1 variants displayed two bands separated by 5-10 kDa. No apparent differences in size or signal intensity were observed amongst CEL-HYB1 variants.

Three independent experiments were performed. Quantification of the immunoblots is shown in **Figure 19B**. The graph for the medium fraction shows that CEL-WT had the highest intensity, while all CEL-HYB1 variants exhibited weak signals. For the lysate fraction, CEL-WT and all CEL-HYB1 variants had similar signal intensities. Finally, for the pellet fraction, the CEL-HYB1 variants were more abundant than CEL-WT. CEL-WT was present only as a very weak band. For the medium fraction CEL-WT displayed statistically significant differences with all CEL-HYB1 variants. For the lysate fraction no differences were observed between CEL-WT and CEL-HYB1 variants. Finally, for the pellet fractions the differences were statistically significant between CEL-WT and, CEL-HYB1 and CEL-HYB1-C587A. Also CEL-HYB1-C583A and CEL-HYB1-C583A/C587A appeared different from CEL-WT, but the inter-experimental variation was probably too large for significance to be reached.

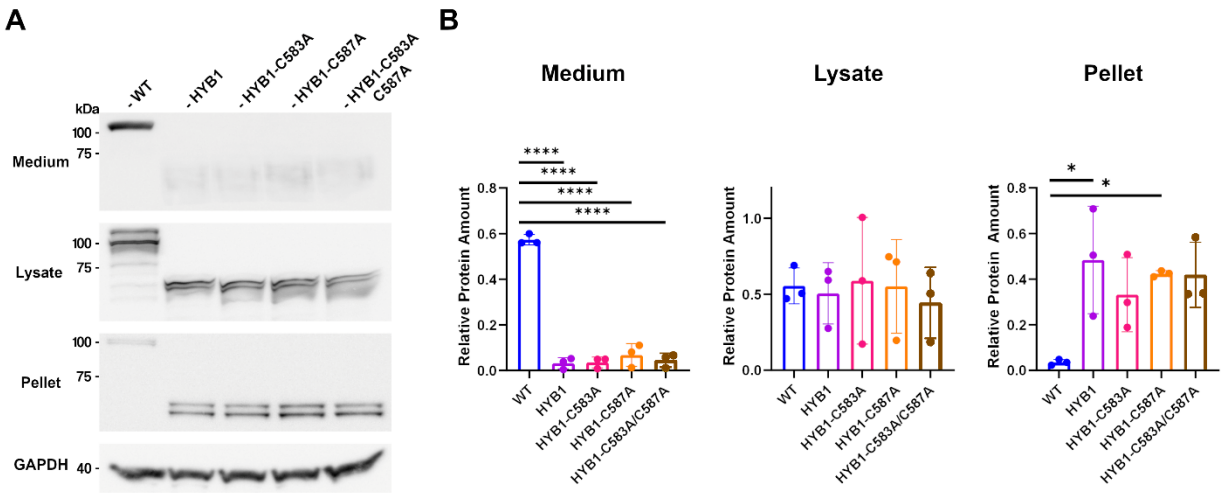

Figure 19. Secretion and intracellular protein distribution of CEL variants in transfected HEK293 cells. (A) Cells were transfected with the CEL-WT, CEL-HYB1, CEL-HYB1-C583A, CEL-HYB1-C587A and CEL-HYB1-C583A/C587A constructs. SDS-PAGE and western blot were used to analyse the medium, lysate and pellet fractions using Anti-V5 antibody. Blots shown are representative of three independent experiments. GAPDH was used as loading control (B)) Quantification of chemiluminiscence intensity of the ER stress markers; three independent experiments were performed and GAPDH signal was used for normalizing. Error bars represent standard deviation (SD). Statistical significance was assessed using one-way ANOVA analysis where * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$.

5.3.2 Endoplasmic reticulum stress markers analysis via western blot

Similar to **Section 5.2.3**, CEL-HYB1 variants were evaluated to investigate if changes in physicochemical properties caused by cysteines can induce ER stress. Pellet fractions isolated in **Section 5.3.1** were analysed via SDS-PAGE and western blotting (**Figure 20A**). Immunodetection of Calnexin, BiP and PERK was performed using specific antibodies against each protein in conjunction with HRP-conjugated secondary antibody. GAPDH was analysed in the lysate fraction and included as a loading control. Calnexin and BiP were located at their expected molecular weights of 75 kDa. However, PERK signal was expected at 140 kDa but no signal was detected, probably caused by defective transfer to the membrane.

Three independent replicates of each ER stress marker were performed. Marker intensity was quantified, and the results are presented in **Figure 20B**. For both calnexin and BiP no statistically significant differences were observed between the different variants. The largest difference was noted between CEL-WT and CEL-HYB1, but it did not reach statistical significance.

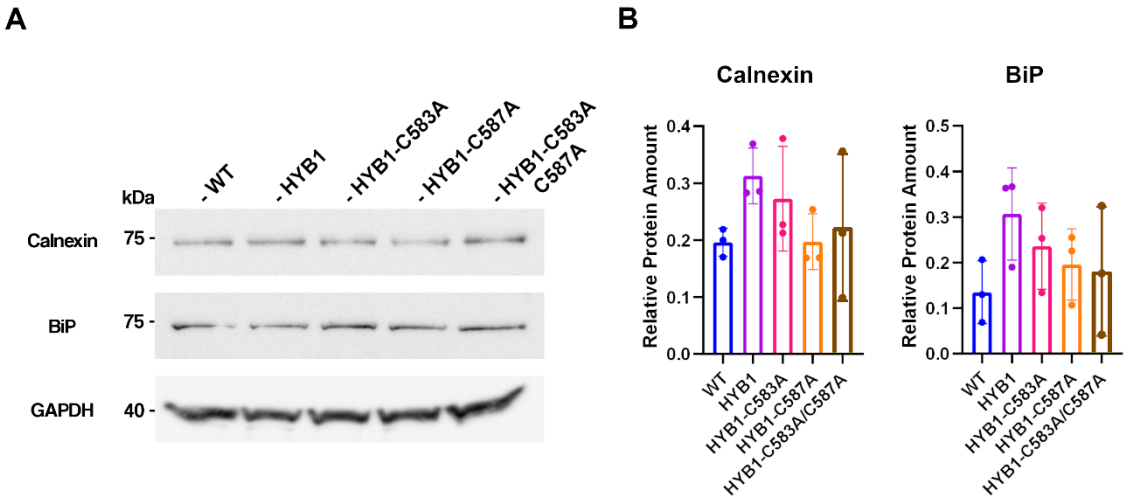

Figure 20. Protein abundance of endoplasmic reticulum stress markers in transfected HEK293 cells. (A) Cells were transfected with the CEL-WT, CEL-HYB1, CEL-HYB1-C583A, CEL-HYB1-C587A and CEL-HYB1-C583A/C587A constructs. SDS-PAGE and western blot were used to analyse pellet fractions. Blots shown are representative of three independent experiments. GAPDH was measured in the lysate fraction and used as loading control (B) Quantification of. ER stress markers intensity for the three independent experiments was normalized according to the GAPDH signal. Error bars represent standard deviation (SD). Statistical significance was assessed using one-way ANOVA analysis where * $P \leq 0.05$.

5.3.3 CEL intracellular protein localization via immunocytochemistry

Similar to the analysis described in **Section 5.2.4,** HEK293 cells transiently transfected with CEL-WT, CEL-HYB1, CEL-HYB1-C583A, CEL-HYB1-C587A and CEL-HYB1-C583/C587A constructs, were immunostained to investigate intracellular CEL distribution and its relationship to different ER stress and Golgi markers. Fixed cells were stained for CEL using the Anti-V5 antibody and then co-stained with one of the ER stress markers calnexin and BiP, or the Golgi marker GALNT2. The nuclei were visualized by using DAPI.

Confocal images of the CEL protein variants co-stained with calnexin, BiP or GALNT2 are shown in **Figure 21, Figure 22** and **Figure 23**, respectively. Green signal represents the intracellular distribution of CEL whereas the red signals mark the other proteins. For CEL-WT, the intracellular distribution was again characterized by localized regions resembling the distribution of the Golgi apparatus. All CEL-HYB1 variants displayed a similar distribution, with localized regions resembling the distribution of the Golgi apparatus and some areas presenting a network-like pattern, resembling the distribution of the ER. No apparent differences in intracellular distribution were appreciated amongst CEL-HYB1 variants.

Regarding the ER stress markers, in both calnexin and BiP (**Figure 21** and **Figure 22**) no differences were observed regarding the distribution of these markers inside the cell. Moreover, All CEL-HYB1 variants showed a slightly higher degree of colocalization with these two markers, represented by more yellow colour being present in the merged images. For GALNT2 (**Figure 23**), no differences were observed regarding its subcellular localization. Both CEL-WT and CEL-HYB1 variants colocalized with GALNT2, represented by yellow colour in the merged images.

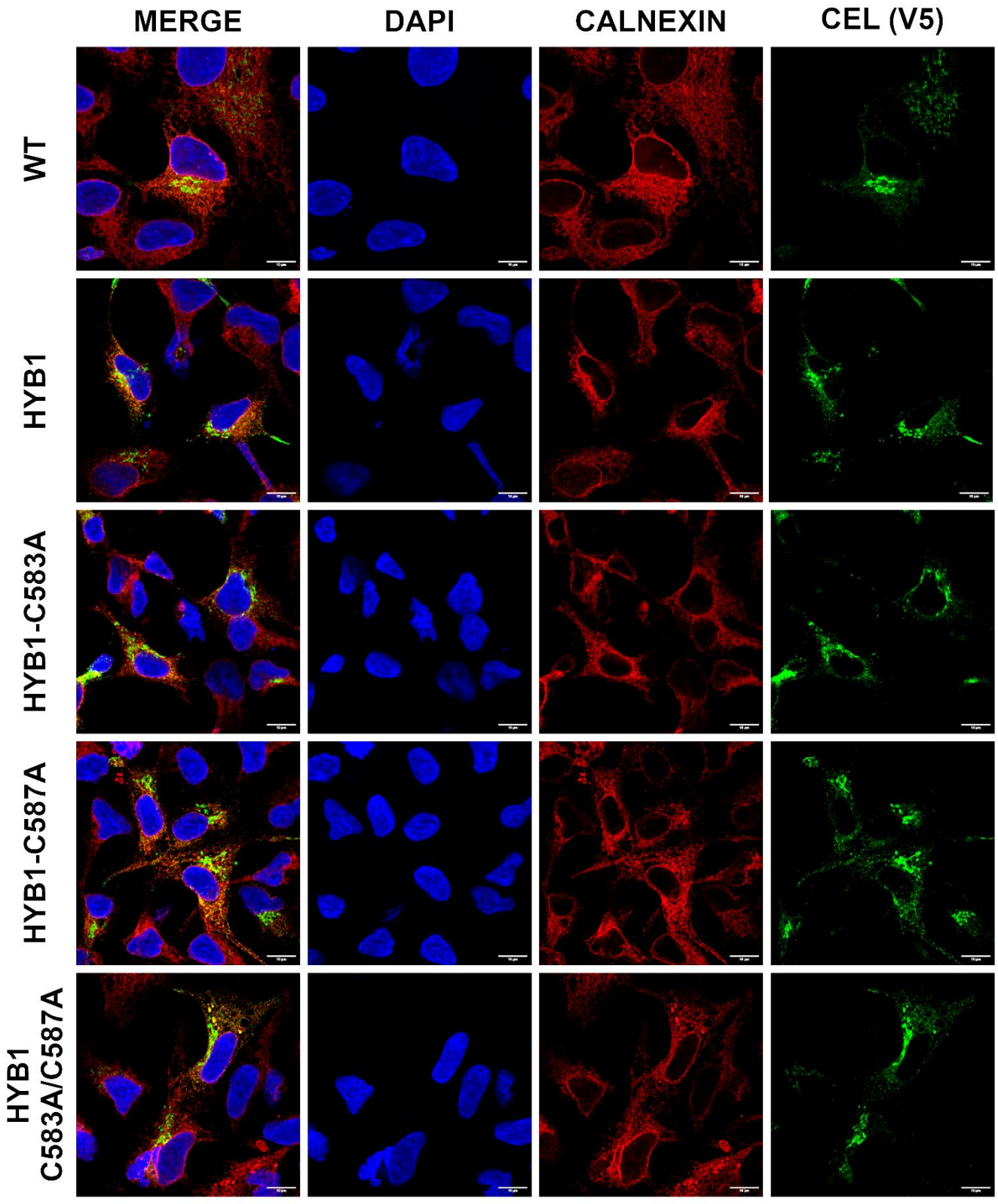

Figure 21. Intracellular distribution of CEL and calnexin in HEK293 cells. Cells were transfected with CEL-WT, CEL-HYB1, CEL-HYB1-C583A, CEL-HYB1-C587A and CEL-HYB1-C583A/C587A, fixed and stained for CEL (Anti-V5, green fluorescence) and Calnexin (red fluorescence). Nuclei are stained blue with DAPI. Images were acquired in a confocal microscope at 100X magnification. Scale bar = 10µm

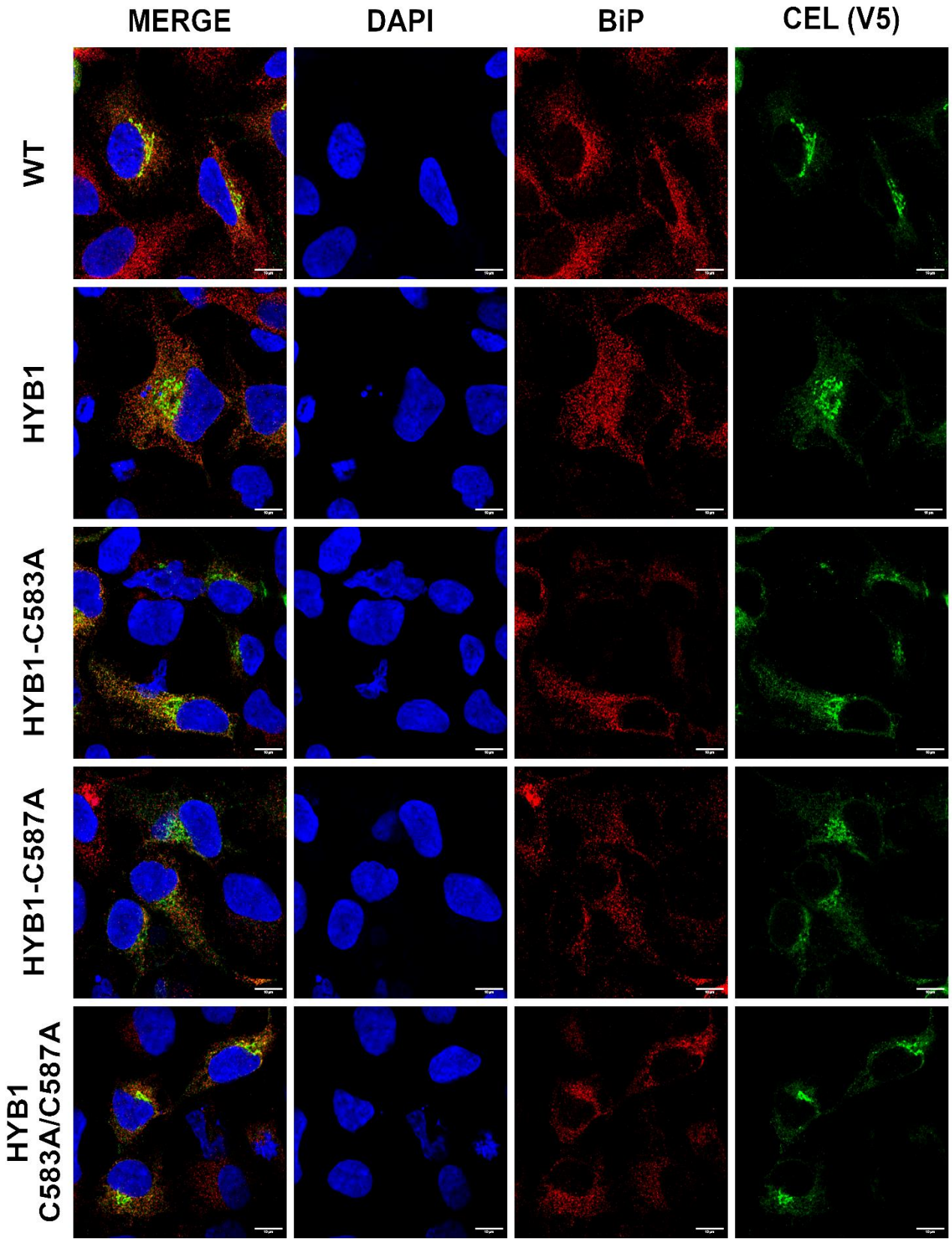

Figure 22. Intracellular distribution of CEL and BiP in HEK293 cells. Cells were transfected with CEL-WT, CEL-HYB1, CEL-HYB1-C583A, CEL-HYB1-C587A and CEL-HYB1-C583A/C587A, fixed and stained for CEL (Anti-V5, green fluorescence) and BiP (red fluorescence). Nuclei are stained blue with DAPI. Images were acquired in a confocal microscope at 100X magnification. Scale bar = 10µm

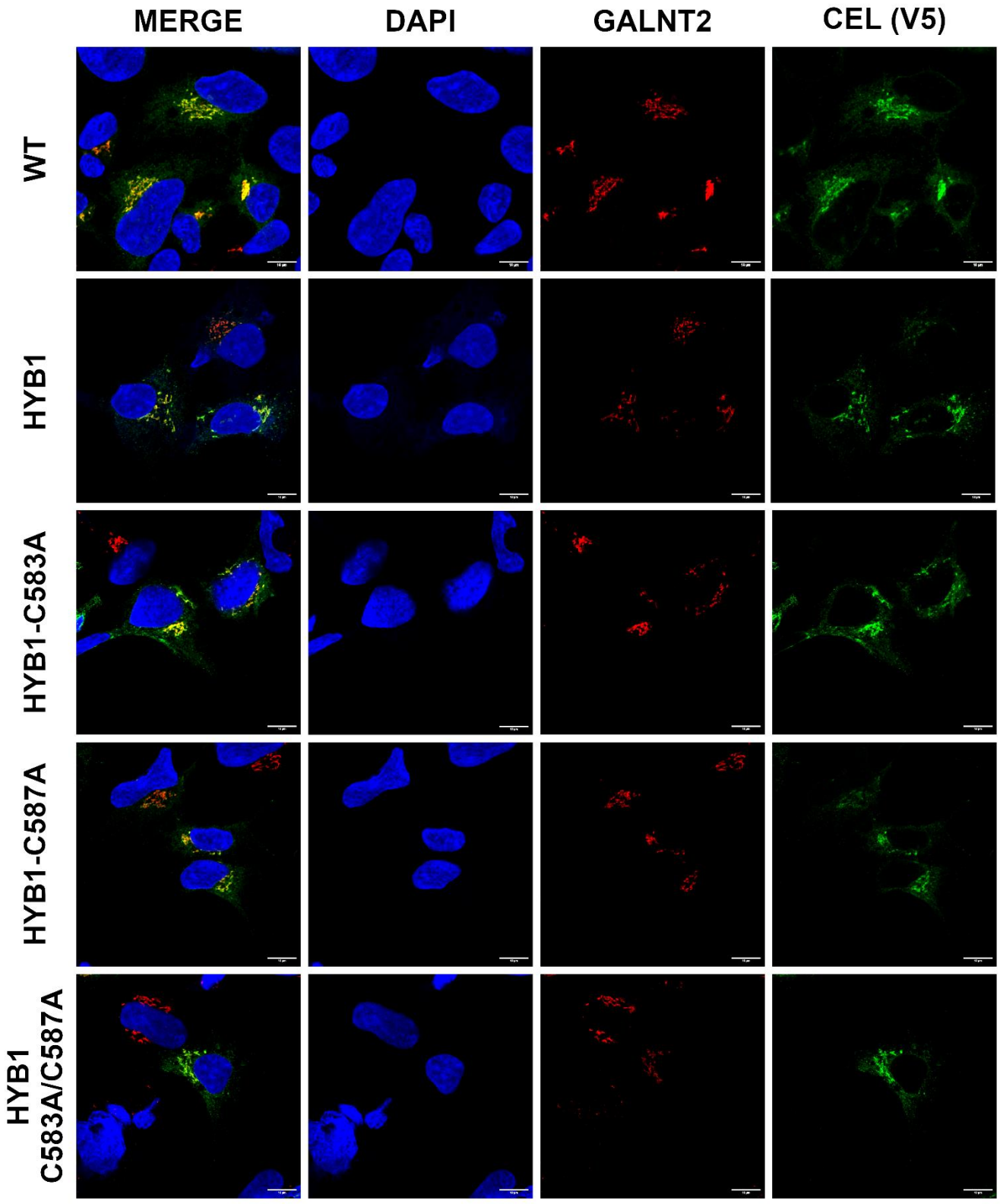

Figure 23. Intracellular distribution of CEL and GALNT2 in HEK293 cells. Cells were transfected with CEL-WT, CEL-HYB1, CEL-HYB1-C583A, CEL-HYB1-C587A and CEL-HYB1-C583A/C587A, fixed and stained for CEL (Anti-V5, green fluorescence) and GALNT2 (red fluorescence). Nuclei are stained blue with DAPI. Images were acquired in a confocal microscope at 100X magnification. Scale bar = 10µm

As done for CEL-MODY variants, the colocalization with calnexin, BiP and GALNT2 was also assessed for CEL-HYB1 variants using the Pearson's coefficient. The coefficients were calculated as described in **Section 5.2.4**. The data from this analysis are presented in **Figure 24**. Although not all differences were significant, the overall picture is that all CEL-HYB1 variants colocalized slightly more than CEL-WT with Calnexin and BiP, therefore being more abundant in the ER. CEL-WT had a significantly higher degree of colocalization with GALNT2 compared to all CEL-HYB1 variants, hence it is more located inside the Golgi. In the colocalization calculations, CEL-HYB1 variants were never significantly different from each other.

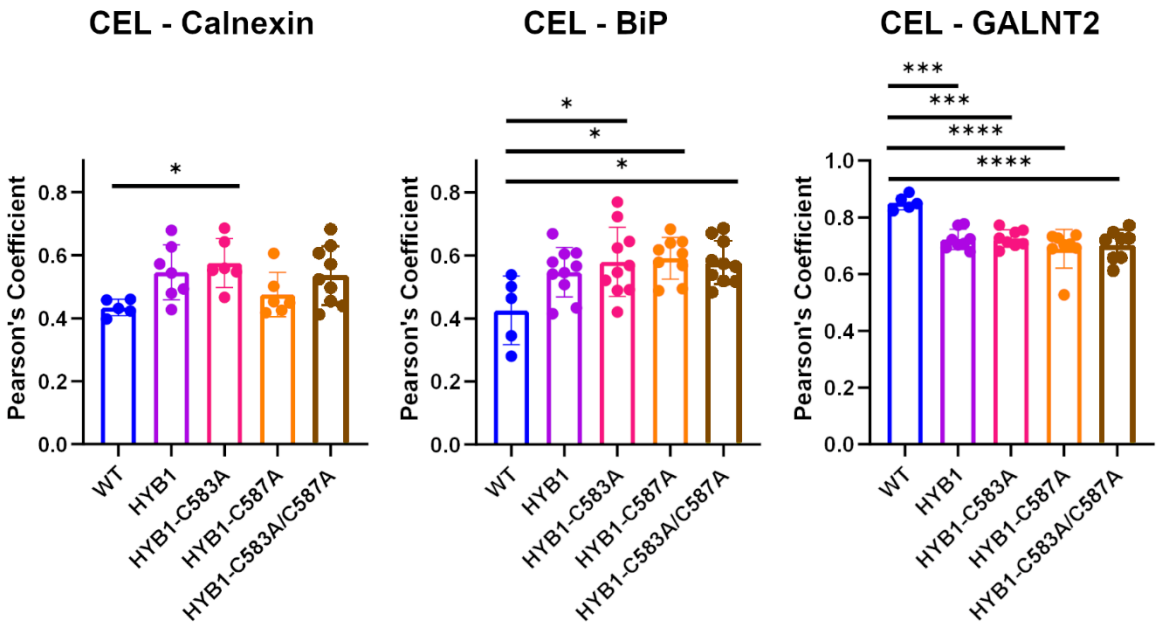

Figure 24. CEL colocalization with ER stress and Golgi markers. Pearson's coefficients were calculated based on immunocytochemistry images from HEK293 cells co-stained for CEL and Calnexin, BiP or GALNT2. Pearson's coefficients were calculated for 5 cells per variant and 4 stacks of each cell were used to calculate the mean. Error bars represent standard deviation (SD). Statistical significance was assessed using one-way ANOVA analysis where ** $P \leq 0.01$, **** $P \leq 0.0001$.

6. Discussion

Both pathogenic CEL variants described so far, CEL-MODY and CEL-HYB1, are characterized by changes in the amino acid sequence of the protein tail encoded by the VNTR region. More specifically, both variants involve the introduction of cysteines, a strong order-promoting amino acid. Acquisition of cysteines often causes retention of mutated proteins in the ER by thiol-mediated mechanisms [132] and aberrant thiol-mediated interactions via unpaired cysteines can directly induce ER retention and aggregation besides misfolding and lead to severe diseases, as for example seen in mutations of FGFR3 or Notch3 [127-131]. Moreover, changes in IDRs leading to the introduction of order-promoting amino acids have been described in human disorders [115]. As described more thoroughly in the Introduction, in CEL-MODY the VNTR has a single-bp deletion leading to a translation frameshift that causes the introduction of ten cysteines in the tail [68, 82]. For CEL-HYB1, recombination between the *CEL* and *CELP* genes causes a truncated tail of only three repeats, but containing two cysteines not presented in the normal CEL protein [80]. Thus, we hypothesized that a main driver of CEL-MODY and CEL-HYB1 pathogenicity is the introduction of cysteine residues in their tail regions. To test this hypothesis, synthetic constructs of the two variants were created, in which the cysteines were mutated into alanine residues, followed by a comparison with the regular constructs as well as the normal CEL protein.

6.1 Secretion and intracellular protein distribution

We started by transiently transfecting HEK293 cells with the constructs of interest and analysed different cellular fractions by western blots. Regarding secretion and intracellular distribution (**Figure 12B**), CEL-MODY-CYS presented statistically significant impaired secretion in the medium when compared to CEL-WT as well as a trend indicating lower levels in the lysate fraction and much higher abundance in the detergent-insoluble fraction. These results corroborate previous studies performed on CEL-MODY-CYS [72, 82, 85, 86, 142]. Intriguingly, for CEL-MODY-ALA the pattern changed completely. This variant behaved similarly to CEL-WT, with increased secretion in the medium fraction and increased presence in the lysate fraction compared to CEL-MODY-CYS. For the pellet fraction, the result with CEL-MODY-ALA was in between what was observed for CEL-WT and CEL-MODY-CYS, with a higher presence than CEL-WT but around half the amount of CEL-MODY-CYS. Thus, the results of the initial study indicate that the cysteines of CEL-MODY are mainly responsible

for its low secretion levels and increased presence in the insoluble pellet fraction. Nevertheless, the behaviour of CEL-MODY-ALA in the pellet fraction, which still was different from CEL-WT, suggests that other aspects of the VNTR-encoded region may play a role for pathogenicity. This may be caused by differences in the number of potential O-glycosylation sites between CEL-WT (n = 36) and CEL-MODY (n = 13) [86], which may influence CEL solubility and stability [143]. In addition, we propose that the addition of 19 arginine residues to the mutated tail [82] may drive phase separation, as arginine-rich frameshifts have been described in hundreds of proteins and usually alter phase separation and partitioning of proteins into biomolecular condensates [115].

Concerning CEL-HYB1 secretion and intracellular distribution (**Figure 19**), when compared to CEL-WT, CEL-HYB1 variants presented statistically significant impaired secretion to the medium fraction, no differences in the lysate and higher presence in the pellet fraction. These results evidence a deficit in CEL-HYB1 secretion as well as intracellular accumulation as detergent-insoluble aggregates in the pellet fraction. Thus, our results support previous knowledge about CEL-HYB1 behaviour [80, 144, 145]. For the ALA-mutated CEL-HYB1 variants, no differences could be observed when compared to CEL-HYB1. Although more experiments are needed to verify this finding, our initial results show that cysteines included in CEL-HYB1 are probably not responsible for the pathogenicity of this variant.

6.2 ER stress

Previous studies have shown that CEL-MODY and CEL-HYB1 increase levels of ER stress markers in the pellet fraction of HEK293 cells [85, 86, 145]. To assess whether cysteines play a role in ER stress development, pellet fractions from HEK293 cells transfected with the CEL variants of interest underwent western blotting to determine calnexin, BiP and PERK levels. For all CEL-MODY and CEL-HYB1 variants, no significant differences for these ER stress markers were observed when compared to CEL-WT (**Figure 13 and Figure 20**). These results are contrasting with what has been described in the literature [85, 86, 145]. However, the differences between studies are probably caused by experimental changes, such as lower transfections rates, leading to dilution of the pathogenic effect, or the use of different reagents. Previous studies used different transfection methods (FuGENE in Xiao *et al*. [85] compared to Lipofectamine 3000 in this thesis), higher DNA amounts (4 µg of DNA in Gravdal *et al.* [86] compared to 2.5 µg of DNA for this thesis) and different antibodies [145]. In addition, PERK was especially difficult to assess as its large molecular weight (140 kDa) made

transfection onto PVDF membranes difficult (as described in the manufacturer's protocol), leading to weak signals in the western blot membranes. Thus, further optimization of the experiment is needed to conclude about the effect of cysteine residues in CEL-MODY and CEL-HYB1 concerning upregulation of UPR proteins.

6.3 Intracellular localization

To further examine the effect of cysteine residues on the CEL protein's distribution, transiently transfected HEK293 cells underwent immunofluorescent staining to assess intracellular localization. When studying CEL-MODY, in line with the results obtained via western blotting, both CEL-WT and CEL-MODY-ALA presented similar behaviour being found in localized regions resembling the cisternae of the Golgi apparatus (**Figure 14, Figure 15 and Figure 16**). Moreover, automatic analysis of colocalization (**Figure 17**), revealed a stronger colocalization of these two variants with GALNT2, an O-glycosylation protein located inside the Golgi apparatus [141]. In contrast, CEL-MODY-CYS showed a diffuse network-like pattern that was reminiscent of the ER. Furthermore, when analysing the colocalization with ER and Golgi markers, CEL-MODY-CYS was more colocalized with calnexin and BiP, chaperones located inside the ER and implicated in the unfolded protein response [139, 140, 146]. This suggests that cysteines in CEL-MODY lead to misfolding of the protein and activation of the unfolded protein response, as already suggested by previous studies [85, 86]. This may be caused by intra and/or extra-molecular disulphide bridges between cysteines that hinder proper folding of CEL.

Furthermore, when examining CEL-HYB1 intracellular distribution, CEL-WT was again found in localized regions resembling the Golgi apparatus and there was a stronger colocalization with GALNT2. In contrast, all tested CEL-HYB1 variants were present in localized regions, in a diffuse network-like pattern reminiscent of the ER (**Figure 21, Figure 22 and Figure 23**), with no apparent differences between the different CEL-HYB1 constructs. In addition, co-localization analysis (**Figure 24**) showed higher colocalization with calnexin and BiP. Our results therefore confirm previous knowledge about CEL-HYB1 [80, 144, 145] and suggest that cysteines in CEL-HYB1 do not seem to play a major role in how this variant behaves.

6.4 Cysteines and CEL-MODY O-linked glycosylation

When examining CEL distribution via western blotting (**Figure 12A**), both CEL-WT and CEL-MODY-ALA exhibited a double band pattern when present in the lysate, while CEL-MODY-CYS had one band and a faint smear. The bands may correspond to O-glycosylated (upper band) and non-O-glycosylated (lower band) versions of the protein, as CEL is highly O-glycosylated in the VNTR region [68]. We intended to examine whether the cysteines in CEL-MODY-CYS could hinder the accessibility of glycosylation machinery. To this end, HEK293 cells and "Simple Cells" (SC) were transfected with CEL-WT, CEL-MODY-CYS and CEL-MODY-ALA. When looking at the western blots (**Figure 18A**) SC presented higher amounts of protein than HEK293 cells. This is probably caused by a higher transfection rate of SC. Cells were seeded in the same quantity and for the same time before transfection, but COSMC knockdown cells have been shown to proliferate less well than wild type HEK293 cells [147]. The result is that there will be fewer cells to transfect with the same number of DNA-plasmid complexes, and therefore higher transfection yield.

Regarding the band patterns (**Figure 18A**), the lysate fraction deriving from SC did not show the characteristic double band pattern for CEL-WT and CEL-MODY-ALA. This indicates that, as suspected, the upper bands were O-glycosylated variants of CEL and that cysteines in CEL-MODY-CYS interfere with the O-glycosylation of the C-terminal tail, probably by limiting the access of O-glycosylation machinery to the glycosylation sites. Furthermore, when analysing protein amounts (**Figure 18B**), the distribution of CEL was maintained in the medium and lysate fractions, but changed for the pellet fraction, with CEL-WT increasing its presence and CEL-MODY-ALA having the same behaviour as CEL-MODY-CYS. This shows that O-glycosylation may have some effect on the behavioural differences of CEL-MODY-CYS and CEL-MODY-ALA, but that it does not account for all variation observed between CEL-WT and CEL-MODY. The differences between CEL-WT and CEL-MODY variants in SC may be explained by the VNTR sequence, as CEL-MODY presents an increased number of arginine residues that have been described to increase phase separation in multiple pathogenic proteins [115]. We suggest that in HEK293 cells the O-Glycosylation helps with reducing the effect of these arginines, as O-glycosylation has been shown to regulate liquid-liquid phase separation [143]. On the other hand, the lack of O-glycosylation in SC may lead to liquid-to-solid phase transition and aggregation of CEL-MODY-ALA [143]. Further analysis of O-glycosylation patterns, like the glycomics experiments performed in El Jellas *et*

al. [79], of the different CEL-MODY variants may be needed to fully understand how cysteine residues affect O-glycosylation and which specific positions are they affecting

6.5 Mechanistic differences of pathogenicity between CEL-MODY and CEL-HYB1

The findings presented above strongly suggest that CEL-MODY and CEL-HYB1 cause pancreatic disease by different mechanisms at the protein level. Regarding CEL-MODY, we propose that cysteines introduced in this protein play a major role in pathogenicity, as removal of cysteines in CEL-MODY-ALA revert properties already reported as common for CEL-MODY (impaired secretion, decreased solubility, reduced O-glycosylation) into properties resembling CEL-WT [68, 72, 85, 86]. Cysteine removal leads to recovery of secretion, reduced presence in the detergent-insoluble pellet fraction, restoring of normal intracellular localization to the Golgi apparatus and increased O-glycosylation. As a mechanism for the pathogenicity, we suggest that the changes in the tail region, specifically the array of cysteines and addition of arginine residues, lead to misfolding of CEL. This causes a loss of solubility and increased colocalization with Calnexin and BiP, both folding chaperones of the ER, leading to ER stress. Furthermore, those protein molecules that manage to fold correctly may then be less glycosylated because of intra or intermolecular disulphide bridges and the reduced number of O-glycosylation sites [86], which may contribute to the lower levels of secretion of this protein. In addition, the positive charges of the added arginine change the *pI* of CEL from the acidic to the basic side [82]. Taking into account that all affected individuals are heterozygotes, the interaction between one negatively charged normal CEL and one positively charged mutant CEL may cause attraction between molecules and promote aggregation.

In contrast, results for CEL-HYB1 indicate a different mechanism where the cysteine residues do not significantly influence this variant's behaviour. We suggest that the mechanistic differences between CEL-HYB1 and CEL-MODY may be caused by the small separation between the two cysteine residues and the short tail of CEL-HYB1. Firstly, the cysteines (C583 and C587) in the CEL-HYB1 tail region are separated only by 3 amino acids. This small distance may make intramolecular disulphide bridges formation difficult as studies have suggested that loop sizes between 10 to 17 amino acids are preferred in such intramolecular bridges [148]. Furthermore, the sequence between these cysteines is Pro-Arg-Pro, amino acids that have large side chains and that may hinder the formation of the loop even more. Moreover,

the short tail length may lead to less accessibility of the tail region, which causes difficulties for intermolecular disulphide bridges formation. Hence, we hypothesize that similar to other short versions of CEL, such as CEL-3R [69], the absence of a long tail may result in pathogenicity due to the loss of solubility provided by the IDR properties of the VNTR-encoded region. Moreover, similar to CEL-MODY, the addition of arginines to the short tail may be related to the increased levels of CEL-HYB1 observed in the pellet fraction as detergent-insoluble aggregates, caused by phase separation [115].

Despite the limitations of the study, we present results strongly suggesting that there are fundamental differences in the underlying mechanisms at the amino acid sequence level between CEL-MODY and CEL-HYB1 pathogenicity. Regardless of these differences, the overall result is the same for both variants, characterized by impaired secretion, intracellular accumulation as detergent-insoluble aggregates and endoplasmic reticulum stress.

6.6 Possible functional role of the *CEL* VNTR

Although the biological importance of the CEL protein tail is still enigmatic, the findings previously described, together with general properties attributed to IDRs may suggest some possible functions. IDRs have been demonstrated to increase solubility [107] or aggregation [108] of proteins, and to modulate phase separation based on their amino acid sequence [109, 110]. Moreover, IDRs commonly undergo post-translational modifications, as the flexibility of these regions facilitates post-translational machinery access to their specific sites [111-113]. The tail region of normal CEL contains a large number of disorder-promoting amino acids, especially the repeated pattern of the PEST sequence [75], which is common for IDRs that increase solubility and mediate the formation of biomolecular condensates [109, 110]. CEL also contains multiple O-glycosylation sites in the VNTR-encoded region [77]. This O-glycosylation may be related to defining the structure of this IDR region as an extended rod-like structure [149] or in the modulation of liquid-liquid phase separation and protein stability [143]. This, together with the observation that cysteines of CEL-MODY lead to low secretion, ER stress upregulation, aggregation and defective O-glycosylation suggest that the CEL tail region increases the protein's solubility and stability when secreted in the mother's milk and/or as part of the pancreatic secretome.

Finally, IDRs have also been related to multifunctional proteins, such as moonlighting proteins [106, 114]. One possibility is that the tail of CEL may aid substrate promiscuity as this lipase can hydrolyse both dietary fat, cholesteryl esters and fat-soluble vitamins [57, 58, 63, 64]. On the other hand, studies have shown that the aberrant CEL-MODY or CEL-HYB1 tail [86, 88], or the removal of the VNTR-encoded segment of CEL has almost no impact on the enzyme's catalytic activity [150, 151]. Thus, in our view the most plausible explanation is to connect the CEL VNTR-encoded tail with protein solubility and stability. Nevertheless, the knowledge about this region is still insufficient and additional studies are required to completely elucidate the role of the repeat region in the CEL C-terminal tail.

6.7 Limitations of the study

Even though our results match already existing data about both CEL-MODY and CEL-HYB1, and strongly suggest differences between these two variants, there are some limitations to this study that one should pay attention to in further studies.

6.7.1 Use of V5-tagged plasmids

All constructs used in the study contain a V5 epitope Tag and His Tag fused to the C-terminal of the recombinant CEL protein. The V5-Tag is a small peptide tag (14 amino acids) that can be used in methods like western blotting or immunocytochemistry [152]. The biggest advantage of using a V5-Tag is the easy detection of the protein by employing an anti-V5 antibody, as commercial antibodies against this tag have a high affinity. The His-Tag is a well-known tag used for protein purification [153]. Although these tags facilitate the purification and analysis of CEL, they have been shown to increase CEL secretion and solubility [86].

6.7.2 Cell line choice

HEK293 cells [134] and "Simple Cells" (SC = HEK293 cells with a disrupted gene encoding the COSMC chaperone) were selected for expression of the proteins, followed by studies using western blotting and immunocytochemistry. We chose these cell lines because they are human, enabling proper post-translational modifications (PTMs) and folding of CEL. In addition, they are secretory cells, an important trait when studying digestive enzymes. Moreover, they do not express CEL endogenously, so all expression of CEL arise from the variants to be studied. However, these cells derive from human embryonic kidney cells, and may not be the most representative context, as for example their secretory pathway is unregulated [154]. Thus, it would be interesting for future experiments to use acinar cells. Unfortunately, no established

pancreatic acinar human cell line is available to date. Available lines are either rat acinar cells (ARJ42J) or mouse acinar cells (266-6), which would introduce biological differences when compared to human cells. Moreover, transfection of mouse 266-6 cells was already tested in our group for western blotting and immunocytochemistry, but these cell line proved challenging to culture and had low transfection efficiency using the conventional Lipofectamine methods.

6.7.3 Transfection efficiency improvements

During optimization for this study, transfection was performed using the lowest quantities recommended by the manufacturer, which thanks to the high affinity and specificity of the anti-V5 antibody and that all CEL was derived only from transfected cells, made immunodetection of CEL easy. However, we did never monitor transfection efficiency, by for example using a plasmid expressing a fluorescent protein. Lacking a proper transfection efficiency monitoring, transfection for the study of ER stress markers was performed under the same conditions used for the study of CEL, probably leading to low transfection rates and dilution of the pathogenic effect. To further optimize detection of ER stress markers by western blotting, transfection efficiency should be achieved around 80-90% of transfected cells in order to ensure that the effect of defective CEL on ER stress is not diluted by the presence of non-transfected cells. Another option would be to use stably transfected cells, where the entire population would express the variants of interest.

6.7.4 Effect on enzyme activity

In this study, protein secretion and distribution were analysed and compared between the different CEL variants. However, none of the experiments gave any information about the effect of cysteine residues on enzyme activity. Nonetheless, previous studies of CEL-MODY and CEL-HYB1 have shown no significant differences in enzymatic activity when compared to CEL-WT [86, 88], therefore suggesting that cysteines do not modify CEL's enzymatic activity.

7. Conclusions

The main purpose of this thesis was to broaden the knowledge about the function and properties of the VNTR-encoded region of CEL. We specifically sought to evaluate, the effect of the cysteine residues present in the two known pathogenic variants of the protein, i.e. CEL-MODY and CEL-HYB1. Based on the previously described findings, the following conclusions can be drawn:

- Proper secretion of CEL-MODY is restored when the cysteine residues of the aberrant tail are mutated into alanine residues.
- The cysteine residues in the CEL-MODY tail region cause aggregation of the protein in the detergent-insoluble pellet fraction. However, also other factors such as the increased number of arginine residues may play a role in this accumulation.
- Mutation of cysteines from CEL-MODY tail region into alanine residues reverts the intracellular localization of CEL from a pathogenic state (endoplasmic reticulum) to the normal state (Golgi apparatus).
- Cysteine residues in CEL-MODY interfere with the O-glycosylation of the VNTR-encoded tail.
- Cysteine residues of the tail region do not account for any of the pathogenic changes that the CEL-HYB1 protein undergoes. Protein secretion, distribution, and intracellular localization remain unchanged when substituting these cysteines with alanine residues.
- CEL-MODY and CEL-HYB1 are therefore likely to involve different pathogenic mechanisms at the protein level. For CEL-MODY a main driver of pathogenesis is probably the cysteine residues present in the aberrant protein tail. In contrast, CEL-HYB1 pathogenicity seems to be unaffected by cysteines in its tail region.

8. Future perspectives

The amino acid repeats of the CEL tail region are likely to play a major role in the correct functioning of the protein. This thesis has revealed some important mechanistic differences between the two known pathogenic variants of CEL. Despite the gained knowledge, further studies are needed to refine how cysteines, arginine-rich frameshift and lack of O-glycosylation interact in driving CEL-MODY pathogenicity and how important each of these factors is in this variant's behaviour. Moreover, to decipher the mechanism behind CEL-HYB pathogenicity, an approach directed towards the effect of tail length and the tail arginine residues may be needed. We propose that further experiments should focus on the following:

CEL-MODY disease mechanism:

- Further optimization of the ER stress markers used in the western blot.
- Analysis and comparison of the O-glycome of the MODY-CYS and ALA variants via MALDI-TOF MS.
- Immunocytochemistry of transfected acinar cells to analyse the different variants in a more biologically representative context.
- Protein purification and in vitro droplet formation experiment [115] to assess the role of cysteine residues in phase separation.

CEL-HYB1 disease mechanism:

- Analyse the behaviour of CEL-HYB1 variants in acinar cells so the context is more biologically representative.
- Generate variants containing different sizes of the normal *CEL* VNTR to analyse the effect of tail size on CEL function and pathogenicity.
- Generate mutants based on the removal of arginine residues to address the effect of these amino acids in CEL-HYB1 pathogenicity.

References

1. Reisner, H.M., *Chapter 12: Pathology of the Pancreas. In Lange (Ed.), Pathology: A Modern Case Study*. 2020.
2. Frantz, E., V. Souza-Mello, and C. Mandarim-de-Lacerda, *Pancreas: Anatomy, diseases and health implications.* 2012.
3. Smith, M.E. and D.G. Morton, *The Digestive System: Basic science and clinical conditions (2nd ed.)*. Systems of the body. 2010: Churchill Livingstone Elsevier.
4. Logsdon, C.D. and B. Ji, *The role of protein synthesis and digestive enzymes in acinar cell injury.* Nat Rev Gastroenterol Hepatol, 2013. **10**(6): p. 362-70.
5. Atkinson, M.A., et al., *Organisation of the human pancreas in health and in diabetes.* Diabetologia, 2020. **63**(10): p. 1966-1973.
6. Mayo Clinic. *The pancreas in the digestive system*. [cited 2023 May 2023]; Available from: www.mayoclinic.org/diseases-conditions/pancreatitis/multimedia/pancreas-in-the-digestive-system/img-20005694
7. Kim, A., et al., *Islet architecture: A comparative study.* Islets, 2009. **1**(2): p. 129-36.
8. Da Silva Xavier, G., *The Cells of the Islets of Langerhans.* J Clin Med, 2018. **7**(3).
9. Campbell, J.E. and C.B. Newgard, *Mechanisms controlling pancreatic islet cell function in insulin secretion.* Nat Rev Mol Cell Biol, 2021. **22**(2): p. 142-158.
10. Steiner, D.J., et al., *Pancreatic islet plasticity: interspecies comparison of islet architecture and composition.* Islets, 2010. **2**(3): p. 135-45.
11. Gerich, J.E. and P.J. Campbell, *Overview of counterregulation and its abnormalities in diabetes mellitus and other conditions.* Diabetes Metab Rev, 1988. **4**(2): p. 93-111.
12. *Delta Cell Anatomy Ancyclopaedia Britanicca Inc*. [cited 2023 22/02/2023]; Available from: https://www.britannica.com/science/digestion-biology.
13. Napolitano, T., et al., *Role of ghrelin in pancreatic development and function.* Diabetes Obes Metab, 2018. **20 Suppl 2**: p. 3-10.
14. *Islet of Langerhans anatomy and cell types*. [cited 2023 May, 30th, 2023]; Available from: https://www.pinterest.es/pin/447263806746003042/.
15. Pandiri, A.R., *Overview of exocrine pancreatic pathobiology.* Toxicol Pathol, 2014. **42**(1): p. 207-16.
16. Bolender, R.P., *Stereological analysis of the guinea pig pancreas. I. Analytical model and quantitative description of nonstimulated pancreatic exocrine cells.* J Cell Biol, 1974. **61**(2): p. 269-87.
17. Geron, E., E.D. Schejter, and B.Z. Shilo, *Assessing the secretory capacity of pancreatic acinar cells.* J Vis Exp, 2014(90).
18. Williams, J.A., *Regulation of acinar cell function in the pancreas.* Curr Opin Gastroenterol, 2010. **26**(5): p. 478-83.
19. Leung, P.S. and S.P. Ip, *Pancreatic acinar cell: its role in acute pancreatitis.* Int J Biochem Cell Biol, 2006. **38**(7): p. 1024-30.
20. Zeng, M., L. Vachel, and S. Muallem, *Physiology of Duct Cell Secretion*, in *The Pancreas: an integrated textbook of basic science*, H.G. Beger, et al., Editors. 2018, John Wiley & Sons Ltd. p. 56-62.
21. Federation, I.D., *IDF Diabetes Atlas, 10th edn*. 2021: International Diabetes Federation.
22. Collaboration, N.C.D.R.F., *Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants.* Lancet, 2016. **387**(10027): p. 1513-1530.
23. Seuring, T., O. Archangelidi, and M. Suhrcke, *The Economic Costs of Type 2 Diabetes: A Global Systematic Review.* Pharmacoeconomics, 2015. **33**(8): p. 811-31.
24. Atkinson, M.A., G.S. Eisenbarth, and A.W. Michels, *Type 1 diabetes.* Lancet, 2014. **383**(9911): p. 69-82.
25. Acharjee, S., et al., *Understanding type 1 diabetes: etiology and models.* Can J Diabetes, 2013. **37**(4): p. 269-276.
26. Chatterjee, S., K. Khunti, and M.J. Davies, *Type 2 diabetes.* Lancet, 2017. **389**(10085): p. 2239-2251.
27. Wong, E., et al., *Diabetes and risk of physical disability in adults: a systematic review and meta-analysis.* Lancet Diabetes Endocrinol, 2013. **1**(2): p. 106-14.
28. Fox, C.S., et al., *Trends in cardiovascular complications of diabetes.* JAMA, 2004. **292**(20): p. 2495-2499.
29. Retnakaran, R., et al., *Risk factors for renal dysfunction in type 2 diabetes: U.K. Prospective Diabetes Study 74.* Diabetes, 2006. **55**(6): p. 1832-9.
30. Kleinberger, J.W., et al., *Monogenic diabetes in overweight and obese youth diagnosed with type 2 diabetes: the TODAY clinical trial.* Genetics in Medicine 2018. **20**(6): p. 583-590.

31. Lemelman, M.B., L. Letourneau, and S.A.W. Greeley, *Neonatal Diabetes Mellitus: An Update on Diagnosis and Management.* Clin Perinatol, 2018. **45**(1): p. 41-59.
32. Peixoto-Barbosa, R., A.F. Reis, and F.M.A. Giuffrida, *Update on clinical screening of maturity-onset diabetes of the young (MODY).* Diabetol Metab Syndr, 2020. **12**: p. 50.
33. Vaxillaire, M., et al., *A gene for maturity onset diabetes of the young (MODY) maps to chromosome 12q.* Nature Genetics 1995. **9**(4): p. 418-423.
34. Nkonge, K.M., D.K. Nkonge, and T.N. Nkonge, *The epidemiology, molecular pathogenesis, diagnosis, and treatment of maturity-onset diabetes of the young (MODY).* Clin Diabetes Endocrinol, 2020. **6**(1): p. 20.
35. Peery, A.F., et al., *Burden and Cost of Gastrointestinal, Liver, and Pancreatic Diseases in the United States: Update 2018.* Gastroenterology, 2019. **156**(1): p. 254-272 e11.
36. Yadav, D. and A.B. Lowenfels, *The epidemiology of pancreatitis and pancreatic cancer.* Gastroenterology, 2013. **144**(6): p. 1252-61.
37. Mayerle, J., et al., *Genetics, Cell Biology, and Pathophysiology of Pancreatitis.* Gastroenterology, 2019. **156**(7): p. 1951-1968 e1.
38. Xiao, A.Y., et al., *Global incidence and mortality of pancreatic diseases: a systematic review, meta-analysis, and meta-regression of population-based cohort studies.* Lancet Gastroenterol Hepatol, 2016. **1**(1): p. 45-55.
39. *IAP/APA evidence-based guidelines for the management of acute pancreatitis.* Pancreatology, 2013. **13**(4): p. e1-e15.
40. Boxhoorn, L., et al., *Acute pancreatitis.* The Lancet, 2020. **396**(10252): p. 726-734.
41. Roberts, S.E., et al., *The incidence and aetiology of acute pancreatitis across Europe.* Pancreatology, 2017. **17**(2): p. 155-165.
42. Hoffmeister, A., et al., *English language version of the S3-consensus guidelines on chronic pancreatitis: Definition, aetiology, diagnostic examinations, medical, endoscopic and surgical management of chronic pancreatitis.* Z Gastroenterol, 2015. **53**(12): p. 1447-95.
43. Kleeff, J., et al., *Chronic pancreatitis.* Nat Rev Dis Primers, 2017. **3**: p. 17060.
44. Beyer, G., et al., *Chronic pancreatitis.* The Lancet, 2020. **396**(10249): p. 499-512.
45. Whitcomb, D.C., et al., *Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene.* Nat Genet, 1996. **14**(2): p. 141-5.
46. Whitcomb, D.C., et al., *A gene for hereditary pancreatitis maps to chromosome 7q35.* Gastroenterology, 1996. **110**(6): p. 1975-80.
47. Rosendahl, J., et al., *Hereditary chronic pancreatitis.* Orphanet J Rare Dis, 2007. **2**: p. 1.
48. Gillen, S., et al., *Preoperative/neoadjuvant therapy in pancreatic cancer: a systematic review and meta-analysis of response and resection percentages.* PLoS Med, 2010. **7**(4): p. e1000267.
49. Goral, V., *Pancreatic Cancer: Pathogenesis and Diagnosis.* Asian Pac J Cancer Prev, 2015. **16**(14): p. 5619-24.
50. Klein, A.P., *Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors.* Nat Rev Gastroenterol Hepatol, 2021. **18**(7): p. 493-502.
51. Zhao, Z. and W. Liu, *Pancreatic Cancer: A Review of Risk Factors, Diagnosis, and Treatment.* Technol Cancer Res Treat, 2020. **19**: p. 1533033820962117.
52. Kamisawa, T., et al., *Pancreatic cancer.* The Lancet, 2016. **388**(10039): p. 73-85.
53. Postlewait, L.M., et al., *Pancreatic neuroendocrine tumors: Preoperative factors that predict lymph node metastases to guide operative strategy.* J Surg Oncol, 2016. **114**(4): p. 440-5.
54. Sarantis, P., et al., *Pancreatic ductal adenocarcinoma: Treatment hurdles, tumor microenvironment and immunotherapy.* World J Gastrointest Oncol, 2020. **12**(2): p. 173-181.
55. Vincent, A., et al., *Pancreatic cancer.* Lancet, 2011. **378**(9791): p. 607-20.
56. Rozenblum, E., et al., *Tumor-suppressive pathways in pancreatic carcinoma.* Cancer Res, 1997. **57**(9): p. 1731-1734.
57. Blackberg, L., et al., *Bile salt-stimulated lipase in human milk and carboxyl ester hydrolase in pancreatic juice: are they identical enzymes?* FEBS Lett, 1981. **136**(2): p. 284-8.
58. Hui, D.Y. and P.N. Howles, *Carboxyl ester lipase: structure-function relationship and physiological role in lipoprotein metabolism and atherosclerosis.* J Lipid Res, 2002. **43**(12): p. 2017-30.
59. Whitcomb, D.C. and M.E. Lowe, *Human pancreatic digestive enzymes.* Dig Dis Sci, 2007. **52**(1): p. 1-17.
60. Abouakil, N. and D. Lombardo, *Inhibition of human milk bile-salt-dependent lipase by boronic acids. Impfication to the bile salts activator effect.* Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism, 1989. **1004**(2): p. 215-220.
61. Hernell, O. and T. Olivecrona, *Human milk lipases II. Bile salt-stimulated lipase.* Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism, 1974. **369**(2): p. 234-244.

62. Lombardo, D., J. Fauvel, and O. Guy, *Studies on the substrate specificity of a carboxyl ester hydrolase from human pancreatic juice. I. Action on carboxyl esters, glycerides and phospholipids.* Biochimica et Biophysica Acta (BBA) - Enzymology, 1980. **611**(1): p. 136-146.
63. Lombardo, D. and O. Guy, *Studies on the substrate specificity of a carboxyl ester hydrolase from human pancreatic juice. II. Action on cholesterol esters and lipid-soluble vitamin esters.* Biochimica et Biophysica Acta (BBA) - Enzymology, 1980. **611**(1): p. 147-155.
64. Lombardo, D., O. Guy, and C. Figarella, *Purification and characterization of a carboxyl ester hydrolase from human pancreatic juice.* Biochimica et Biophysica Acta (BBA) - Enzymology, 1978. **527**(1): p. 142-149.
65. Ellis, L.A. and M. Hamosh, *Bile salt stimulated lipase: Comparative studies in ferret milk and lactating mammary gland.* Lipids, 1992. **27**(11): p. 917-922.
66. Taylor, A., et al., *Carboxyl ester lipase: A highly polymorphic locus on human chromosome 9qter.* Genomics, 1991. **10**(2): p. 425-431.
67. Lidberg, U., et al., *Genomic Organization, Sequence Analysis, and Chromosomal Localization of the Human Carboxyl Ester Lipase (CEL) Gene and a CEL-Like (CELL) Gene.* Genomics, 1992. **Jul;13(3):630-40**.
68. Raeder, H., et al., *Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction.* Nat Genet, 2006. **38**(1): p. 54-62.
69. Torsvik, J., et al., *Mutations in the VNTR of the carboxyl-ester lipase gene (CEL) are a rare cause of monogenic diabetes.* Hum Genet, 2010. **127**(1): p. 55-64.
70. Madeyski, K., et al., *Structure and organization of the human carboxyl ester lipase locus.* Mammalian Genome, 1998. **9**: p. 334-338
71. Nilsson, J., M. Hellquist, and G. Bjursell, *The Human Carboxyl Ester Lipase-like (CELL) Gene Is Ubiquitously Expressed and Contains a Hypervariable Region.* Genomics, 1993. **17**(2): p. 416-422.
72. Johansson, B.B., et al., *The role of the carboxyl ester lipase (CEL) gene in pancreatic disease.* Pancreatology, 2018. **18**(1): p. 12-19.
73. Terzyan, S., et al., *Crystal structure of the catalytic domain of human bile salt activated lipase.* Protein Sci 2000. **9**(9): p. 1783-1790.
74. Reue, K., et al., *cDNA cloning of carboxyl ester lipase from human pancreas reveals a unique proline-rich repeat unit.* Journal of Lipid Research 1991. **32**(2): p. 267-276.
75. Rogers, S. and R. Wells, *Amino acid sequences common to rapidly fdegraded proteins: the PEST hypothesis.* Science, 1986. **234**(4774): p. 364-368.
76. Abouakil, N., et al., *Bile salt-dependent lipase biosynthesis in rat pancreatic AR 4-2 J cells. Essential requirement of N-linked oligosaccharide for secretion and expression of a fully active enzyme.* Journal of Biological Chemistry, 1993. **268**(34): p. 25755-25763.
77. Bruneau, N., et al., *O-Glycosylation of C-terminal Tandem-repeated Sequences Regulates the Secretion of Rat Pancreatic Bile Salt-dependent Lipase.* Journal of Biological Chemistry, 1997. **272**(43): p. 27353-27361.
78. Pasqualini, E., et al., *Phosphorylation of the rat pancreatic bile-salt-dependent lipase by casein kinase II is essential for secretion.* Biochem J. , 2000. **345**: p. 121-128.
79. El Jellas, K., et al., *The mucinous domain of pancreatic carboxyl-ester lipase (CEL) contains core 1/core 2 O-glycans that can be modified by ABO blood group determinants.* J Biol Chem, 2018. **293**(50): p. 19476-19491.
80. Fjeld, K., et al., *A recombined allele of the lipase gene CEL and its pseudogene CELP confers susceptibility to chronic pancreatitis.* Nat Genet, 2015. **47**(5): p. 518-522.
81. Martinez, E., et al., *Expression of truncated bile salt-dependent lipase variant in pancreatic pre-neoplastic lesions.* Oncotarget, 2017. **8**(1): p. 536-551.
82. Johansson, B.B., et al., *Diabetes and pancreatic exocrine dysfunction due to mutations in the carboxyl ester lipase gene-maturity onset diabetes of the young (CEL-MODY): a protein misfolding disease.* J Biol Chem, 2011. **286**(40): p. 34593-605.
83. Pellegrini, S., et al., *Generation of beta Cells from iPSC of a MODY8 Patient with a Novel Mutation in the Carboxyl Ester Lipase (CEL) Gene.* J Clin Endocrinol Metab, 2021. **106**(5): p. e2322-e2333.
84. El Jellas, K., et al., *Two New Mutations in the CEL Gene Causing Diabetes and Hereditary Pancreatitis: How to Correctly Identify MODY8 Cases.* J Clin Endocrinol Metab, 2022. **107**(4): p. e1455-e1466.
85. Xiao, X., et al., *A Carboxyl Ester Lipase (CEL) Mutant Causes Chronic Pancreatitis by Forming Intracellular Aggregates That Activate Apoptosis.* J Biol Chem, 2016. **291**(44): p. 23224-23236.
86. Gravdal, A., et al., *The position of single-base deletions in the VNTR sequence of the carboxyl ester lipase (CEL) gene determines proteotoxicity.* J Biol Chem, 2021. **296**: p. 100661.
87. Lupski, J.R. and P. Stankiewicz, *Genomic Disorders: The Genomic Basis of Disease*. 2006, Humana Totowa, NJ. p. 341.

88. Cassidy, B.M., et al., *Single nucleotide polymorphisms in CEL-HYB1 increase risk for chronic pancreatitis through proteotoxic misfolding.* Hum Mutat, 2020. **41**(11): p. 1967-1978.
89. Lee, C. and M.H. Yu, *Protein folding and diseases.* J Biochem Mol Biol, 2005. **38**(3): p. 275-80.
90. Hartl, F.U. and M. Hayer-Hartl, *Molecular chaperones in the cytosol: from nascent chain to folded protein.* Science, 2002. **295**(5561): p. 1852-8.
91. Dobson, C.M., *Principles of protein folding, misfolding and aggregation.* Semin Cell Dev Biol, 2004. **15**(1): p. 3-16.
92. Wang, M. and R.J. Kaufman, *Protein misfolding in the endoplasmic reticulum as a conduit to human disease.* Nature, 2016. **529**(7586): p. 326-35.
93. Petersen, T.N., et al., *SignalP 4.0: discriminating signal peptides from transmembrane regions.* Nat Methods, 2011. **8**(10): p. 785-6.
94. Ellis, R.J. and S.M. van der Vies, *Molecular chaperones.* Annu Rev Biochem, 1991. **60**: p. 321-47.
95. Wallis, A.K. and R.B. Freedman, *Assisting oxidative protein folding: how do protein disulphide-isomerases couple conformational and chemical processes in protein folding?* Top Curr Chem, 2013. **328**: p. 1-34.
96. Lang, K., F.X. Schmid, and G. Fischer, *Catalysis of protein folding by prolyl isomerase.* Nature, 1987. **329**(6136): p. 268-70.
97. Liu, C.Y. and R.J. Kaufman, *The unfolded protein response.* J Cell Sci, 2003. **116**(Pt 10): p. 1861-2.
98. Harding, H.P., Y. Zhang, and D. Ron, *Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase.* Nature, 1999. **397**(6716): p. 271-4.
99. Hetz, C. and F.R. Papa, *The Unfolded Protein Response and Cell Fate Control.* Mol Cell, 2018. **69**(2): p. 169-181.
100. Rutkowski, D.T., et al., *Adaptation to ER stress is mediated by differential stabilities of pro-survival and pro-apoptotic mRNAs and proteins.* PLoS Biol, 2006. **4**(11): p. e374.
101. Tabas, I. and D. Ron, *Integrating the mechanisms of apoptosis induced by endoplasmic reticulum stress.* Nat Cell Biol, 2011. **13**(3): p. 184-90.
102. Walter, P. and D. Ron, *The unfolded protein response: from stress pathway to homeostatic regulation.* Science, 2011. **334**(6059): p. 1081-6.
103. Martin, E.W. and A.S. Holehouse, *Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof.* Emerg Top Life Sci, 2020. **4**(3): p. 307-329.
104. Uversky, V.N., C.J. Oldfield, and A.K. Dunker, *Intrinsically disordered proteins in human diseases: introducing the D2 concept.* Annu Rev Biophys, 2008. **37**: p. 215-46.
105. Jones, D.T. and J.J. Ward, *Prediction of disordered regions in proteins from position specific score matrices.* Proteins, 2003. **53 Suppl 6**: p. 573-8.
106. Chakrabarti, P. and D. Chakravarty, *Intrinsically disordered proteins/regions and insight into their biomolecular interactions.* Biophysical Chemistry, 2022. **283**: p. 106769.
107. Oldfield, C.J., et al., *Utilization of protein intrinsic disorder knowledge in structural proteomics.* Biochim Biophys Acta, 2013. **1834**(2): p. 487-98.
108. Oldfield, C.J. and A.K. Dunker, *Intrinsically disordered proteins and intrinsically disordered protein regions.* Annu Rev Biochem, 2014. **83**: p. 553-84.
109. Bhopatkar, A.A., V.N. Uversky, and V. Rangachari, *Disorder and cysteines in proteins: A design for orchestration of conformational see-saw and modulatory functions.* Prog Mol Biol Transl Sci, 2020. **174**: p. 331-373.
110. Banani, S.F., et al., *Biomolecular condensates: organizers of cellular biochemistry.* Nat Rev Mol Cell Biol, 2017. **18**(5): p. 285-298.
111. Borcherds, W., et al., *How do intrinsically disordered protein regions encode a driving force for liquid-liquid phase separation?* Curr Opin Struct Biol, 2021. **67**: p. 41-50.
112. Dunker, A.K., et al., *Intrinsic disorder and protein function.* Biochemistry, 2002. **41**(21): p. 6573-82.
113. Xu, B., et al., *Rigidity enhances a magic-number effect in polymer phase separation.* Nat Commun, 2020. **11**(1): p. 1561.
114. Jeffery, C.J., *Moonlighting proteins--an update.* Mol Biosyst, 2009. **5**(4): p. 345-50.
115. Mensah, M.A., et al., *Aberrant phase separation and nucleolar dysfunction in rare genetic diseases.* Nature, 2023. **614**(7948): p. 564-571.
116. Coskuner-Weber, O., O. Mirzanli, and V.N. Uversky, *Intrinsically disordered proteins and proteins with intrinsically disordered regions in neurodegenerative diseases.* Biophys Rev, 2022. **14**(3): p. 679-707.
117. Watzlawik, J., et al., *Prion protein helix1 promotes aggregation but is not converted into beta-sheet.* J Biol Chem, 2006. **281**(40): p. 30242-50.
118. Wildegger, G., S. Liemann, and R. Glockshuber, *Extremely rapid folding of the C-terminal domain of the prion protein without kinetic intermediates.* Nat Struct Biol, 1999. **6**(6): p. 550-3.

119. Cooper, G.J., et al., *Purification and characterization of a peptide from amyloid-rich pancreases of type 2 diabetic patients.* Proc Natl Acad Sci U S A, 1987. **84**(23): p. 8628-32.
120. Goldsbury, C., et al., *Amyloid fibril formation from full-length and fragments of amylin.* J Struct Biol, 2000. **130**(2-3): p. 352-62.
121. Jaikaran, E.T., et al., *Identification of a novel human islet amyloid polypeptide beta-sheet domain and factors influencing fibrillogenesis.* J Mol Biol, 2001. **308**(3): p. 515-25.
122. Morand, S., et al., *Targeting of the dual oxidase 2 N-terminal region to the plasma membrane.* J Biol Chem, 2004. **279**(29): p. 30244-51.
123. Shergalis, A.G., et al., *Role of the ERO1-PDI interaction in oxidative protein folding and disease.* Pharmacol Ther, 2020. **210**: p. 107525.
124. Borges, C.R. and D.F. Lake, *Oxidative protein folding: nature's knotty challenge.* Antioxid Redox Signal, 2014. **21**(3): p. 392-5.
125. Braakman, I. and D.N. Hebert, *Protein folding in the endoplasmic reticulum.* Cold Spring Harb Perspect Biol, 2013. **5**(5): p. a013201.
126. Sandor, M., et al., *Novel p.G250A Mutation Associated with Chronic Pancreatitis Highlights Misfolding-Prone Region in Carboxypeptidase A1 (CPA1).* Int J Mol Sci, 2022. **23**(24).
127. You, M., et al., *Effect of Pathogenic Cysteine Mutations on FGFR3 Transmembrane Domain Dimerization in Detergents and Lipid Bilayers.* Biochemistry, 2007. **46**(39): p. 11039-11046.
128. Coto, E., et al., *A new de novo Notch3 mutation causing CADASIL.* Eur J Neurol, 2006. **13**(6): p. 628-31.
129. Joutel, A., et al., *De novo mutation in the Notch3 gene causing CADASIL.* Ann Neurol, 2000. **47**(3): p. 388-91.
130. Opherk, C., et al., *CADASIL mutations enhance spontaneous multimerization of NOTCH3.* Hum Mol Genet, 2009. **18**(15): p. 2761-7.
131. Fra, A., E.D. Yoboue, and R. Sitia, *Cysteines as Redox Molecular Switches and Targets of Disease.* Front Mol Neurosci, 2017. **10**: p. 167.
132. Anelli, T., S. Sannino, and R. Sitia, *Proteostasis and "redoxtasis" in the secretory pathway: Tales of tails from ERp44 and immunoglobulins.* Free Radic Biol Med, 2015. **83**: p. 323-30.
133. Steentoft, C., et al., *Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology.* EMBO J, 2013. **32**(10): p. 1478-88.
134. Graham, F.L., et al., *Characteristics of a human cell line transformed by DNA from human adenovirus type 5.* J Gen Virol, 1977. **36**(1): p. 59-74.
135. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.
136. PTBIOP, BIOP JACoP User Manual. 2020. Available from: https://c4science.ch/w/bioimaging_and_optics_platform_biop/image-processing/imagej_tools/jacop_b/
137. The GIMP Development Team, GIMP. 2019. Available from: https://www.gimp.org
138. GraphPad Software, GraphPad Prism. 2023. Available from: https://www.graphpad.com/features
139. Paskevicius, T., et al., *Calnexin, More Than Just a Molecular Chaperone.* Cells, 2023. **12**(3).
140. Bertolotti, A., et al., *Dynamic interaction of BiP and ER stress transducers in the unfoldedprotein response.* Nature Cell Biology 2000. **2**: p. 326-332.
141. Schjoldager, K.T. and H. Clausen, *Site-specific protein O-glycosylation modulates proprotein processing - deciphering specific functions of the large polypeptide GalNAc-transferase gene family.* Biochim Biophys Acta, 2012. **1820**(12): p. 2079-94.
142. Torsvik, J., et al., *Endocytosis of secreted carboxyl ester lipase in a syndrome of diabetes and pancreatic exocrine dysfunction.* J Biol Chem, 2014. **289**(42): p. 29097-111.
143. Li, X., et al., *Emerging roles of O-glycosylation in regulating protein aggregation, phase separation, and functions.* Curr Opin Chem Biol, 2023. **75**: p. 102314.
144. Fjeld, K., et al., *The genetic risk factor CEL-HYB1 causes proteotoxicity and chronic pancreatitis in mice.* Pancreatology, 2022. **22**(8): p. 1099-1111.
145. Tjora, E., et al., *Protein misfolding in combination with other risk factors in CEL-HYB1-mediated chronic pancreatitis.* Eur J Gastroenterol Hepatol, 2021. **33**(6): p. 839-843.
146. Kopp, M.C., et al., *UPR proteins IRE1 and PERK switch BiP from chaperone to ER stress sensor.* Nat Struct Mol Biol, 2019. **26**(11): p. 1053-1062.
147. Hofmann, B.T., et al., *COSMC knockdown mediated aberrant O-glycosylation promotes oncogenic properties in pancreatic cancer.* Mol Cancer, 2015. **14**: p. 109.
148. Hartig, G.R., T.T. Tran, and M.L. Smythe, *Intramolecular disulphide bond arrangements in nonhomologous proteins.* Protein Sci, 2005. **14**(2): p. 474-82.

149. Shogren, R., T.A. Gerken, and N. Jentoft, *Role of glycosylation on the conformation and chain dimensions of O-linked glycoproteins: light-scattering studies of ovine submaxillary mucin.* Biochemistry, 1989. **28**(13): p. 5525-36.
150. Hansson, L., et al., *Recombinant human milk bile salt-stimulated lipase. Catalytic activity is retained in the absence of glycosylation and the unique proline-rich repeats.* J Biol Chem, 1993. **268**(35): p. 26692-8.
151. Downs, D., et al., *Proline-rich domain and glycosylation are not essential for the enzymic activity of bile salt-activated lipase. Kinetic studies of T-BAL, a truncated form of the enzyme, expressed in Escherichia coli.* Biochemistry, 1994. **33**(26): p. 7979-85.
152. Hanke, T., P. Szawlowski, and R.E. Randall, *Construction of solid matrix-antibody-antigen complexes containing simian immunodeficiency virus p27 using tag-specific monoclonal antibody and tag-linked antigen.* J Gen Virol, 1992. **73 (Pt 3)**: p. 653-60.
153. Arnold, F.H., *Metal-affinity separations: a new dimension in protein processing.* Biotechnology (N Y), 1991. **9**(2): p. 151-6.
154. Tausk, F.A., et al., *Expression of a peptide processing enzyme in cultured cells: truncation mutants reveal a routing domain.* Mol Endocrinol, 1992. **6**(12): p. 2185-96.

Appendix A

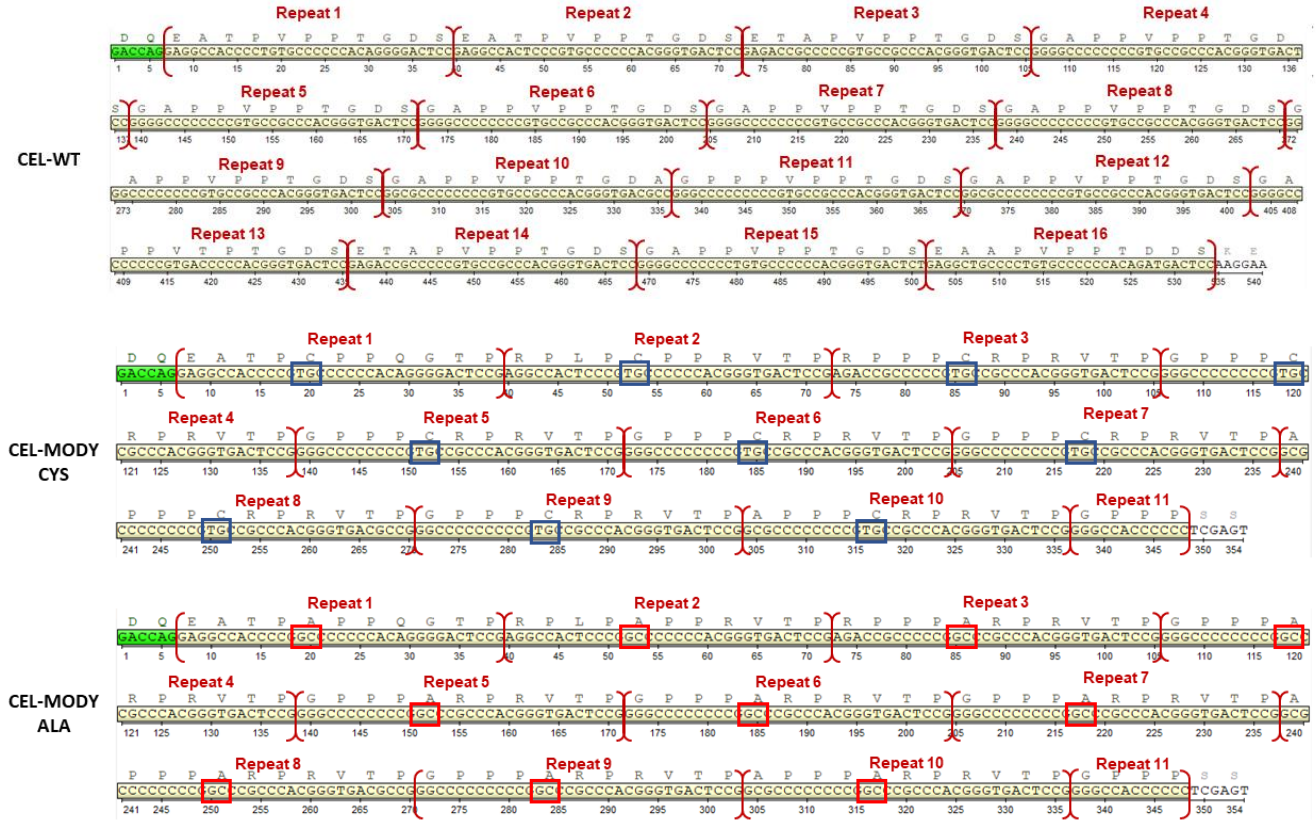

Figure A1. Sanger DNA sequencing results of the CEL-WT, CEL-MODY-CYS and CEL-MODY-ALA constructs VNTR region. Constructs provided by the research group of Prof. Mark Lowe, Washington University School of Medicine, St. Louis, USA. Sequences of CEL-MODY-CYS and CEL-MODY-ALA are highlighted in the codons that were modified. Blue squares indicate the two nucleotides of the TGC codons (coding for cysteines) that were changed into GC, leading to GCC codons that encode alanine. Red squares indicate the two nucleotides that differ from TGC codons leading to GCC codons (coding for alanines) in CEL-MODY-Ala.

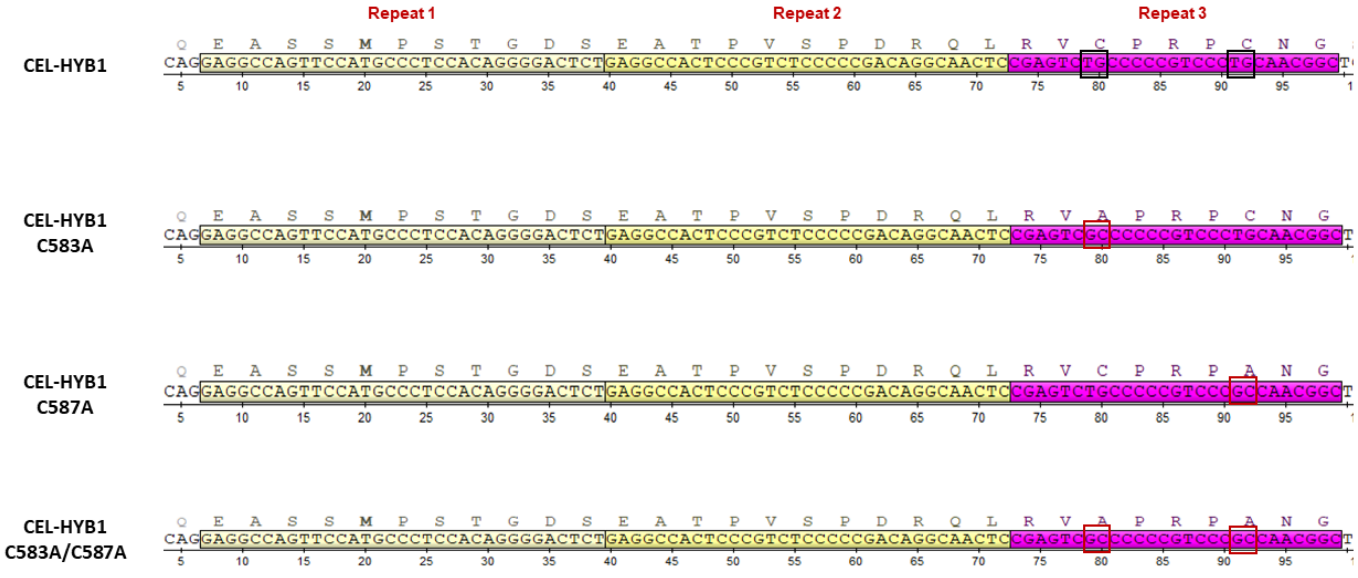

Figure A2. Sanger sequencing results of the different CEL-HYB1 variants. Constructs provided by Khadija El Jellas. Black squares indicate the two nucleotides of the TGC codons (coding for cysteines) that were changed into GC, leading to GCC codons that encode alanine. Red squares indicate the two nucleotides that differ from TGC codons leading to GCC codons (coding for alanines) in CEL-MODY-Ala.