Characterization of insertion variants of the carboxyl-ester lipase (*CEL*) gene - A role in pancreatic disease?

Ranveig Seim Brekke

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

Department of Biological Sciences,

University of Bergen

&

Gade Laboratory for Pathology,

Department of Clinical Medicine

University of Bergen

June 2018

Table of contents

Acknowledgements

The work presented in this thesis was carried out during the period of August 2017 to May 2018 primarily at Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen and at Department of Medical Genetics, Haukeland University Hospital.

First of all I would like to thank my supervisor Anders Molven for giving me the opportunity to write my master thesis about carboxyl ester lipase (CEL). I am grateful for the support and advice during this period and for the counseling throughout the writing process.

My sincere thanks also go to my co-supervisors Karianne Fjeld and Bente Berg Johansson. Karianne, I am grateful for your support, helpful discussions, and wonderful advices during the writing process. I would also like to thank you for the technical help you have given me. Bente, your help and assistance during the writing process, along with your insightful comments and valuable thoughts have been greatly appreciated.

I would also like to thank Solrun Steine and Monika Ringdal. Thank you Solrun for all the technical help and support you have given me throughout the year. Monika, thank you for all the help with reading sequences, I would not have managed to read them without your help.

I would also thank to Anny Gravdal and Khadija EL Jellas. Anny, I thank you for the technical help and your advice during this period. Thank you, Khadija, for the technical help and comments during the experimental work with my thesis.

Lastly, I would like to thank my friends and family for always believing in me and motivating me. A special thanks goes to Sigurd for always being supportive.

Bergen, 2018

Ranveig Seim Brekke

Abbreviations

AA	Amino acid
CEL/*CEL*	Carboxyl ester lipase protein/gene
CELP	Carboxyl ester lipase pseudogene
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EV	Empty vector
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
ICC	Immunocytochemistry
HEK293	Human embryonic kidney cell line 293
IHC	Immunohistochemistry
INS	Insertion of a cytosine in the *CEL* VNTR region
INS9	Insertion in the 9th repeat of the VNTR in the *CEL* gene
INS11	Insertion in the 11th repeat of the VNTR in the *CEL* gene
kDa	Kilo Daltons
mL	Milliliter
o/n	Overnight
PCR	Polymerase chain reaction
PDAC	Pancreatic ductal adenocarcinoma
PP	Pancreatic polypeptide
RT	Room temperature
Rpm	Revolutions per minute
VIP	Vasoactive intestinal peptide
TRUNC	Truncated artificial variant of the *CEL* gene (lacking all of the VNTR region)
VNTR	Variable number of tandem repeats
WT	Wild-type
µg	Microgram
µl	Microliter

Abstract

Carboxyl ester lipase (CEL) is a digestive enzyme that is mainly expressed in the acinar cells of the pancreas. The *CEL* gene is highly polymorphic due to a variable number of tandem repeat (VNTR) region in the last exon. Diseases that are associated with CEL are related to alterations in the VNTR region. Single base insertions in the VNTR cause a premature stop codon and a truncated protein. It is not clear whether these insertions are linked to pancreatic disease.

In this study, we wanted to investigate the location of the insertions and the frequency by sequencing human DNA samples. We further aimed to test tail-specific antibodies for the normal CEL protein (CEL-WT) and the CEL insertion variants (CEL-INS9 and CEL-INS11) by western blot, immunocytochemistry and immunohistochemistry. Moreover, we wanted to study the impact of CEL insertion variants in a cellular model system, thereby evaluating if insertions in the CEL VNTR potentially could be pathogenic.

We sequenced 50 human DNA samples from the Pancreas Biobank at Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen. We noticed five cases with INS9, one case with INS10 and one case with INS12, which results in a carrier frequency of 14 %. However, we observed that the five cases with insertion in repeat 9 all had a VNTR length of 13 on one of the alleles. We further investigated this by sequencing all remaining DNA samples in the Pancreas Biobank with a 13 repeat VNTR length (N=25) and observed that in total, 23 of 31 samples had insertion in repeat 9 and a VNTR length of 13 on one of the alleles.

The CEL-WT protein has a C-terminal that ends with the six amino acids, PAVIRF. However, if an insertion occurs in the CEL VNTR it results in a truncated protein and a C-terminal that ends with the sequence PRAAHG. Tail-specific antibodies have been produced towards these C-terminal ends. We tested the tail-specific antibodies by western blotting, immunocytochemistry and immunohistochemistry. The tail-specific antibodies were highly specific in the western blots, but did not display the same specificity by immunocytochemistry and immunohistochemistry.

Finally, we studied the effect of insertions in the CEL VNTR on protein secretion and aggregation in HEK293 cells. We observed that the insertion variants were similarly detected

in the lysate and the medium. However, we detected more of the CEL-INS9 variant, than CEL-INS11 and CEL-WT in the pellet fraction.

In summary, we have shown that there is an association between insertion in repeat 9 of the *CEL* VNTR and a VNTR length of 13. We also found that the tail-specific antibodies had the highest specificity in western blot analysis. Moreover, our data indicated that insertions may affect the distribution of the CEL protein in different cellular fractions, thereby implicating that the VNTR region is important for the cellular properties of CEL.

1. Introduction

1.1 The pancreas and its diseases

The pancreas is located in the upper part of the abdominal cavity (Fig. 1.1 A), has an elongated structure and is approximately 15-20 cm long (Holck, 2017). Ligaments and vessels demarcate the organ into a head, body and tail region (Kumar et al., 2013).The pancreas consists of an endocrine and an exocrine part. The endocrine part represents 1-2 % of the total volume and produces hormones, which are secreted into the blood stream. The exocrine part, which constitutes the major volume of the organ, excretes digestive enzymes to the duodenum (Fig.1.1B) (Kumar et al., 2013).

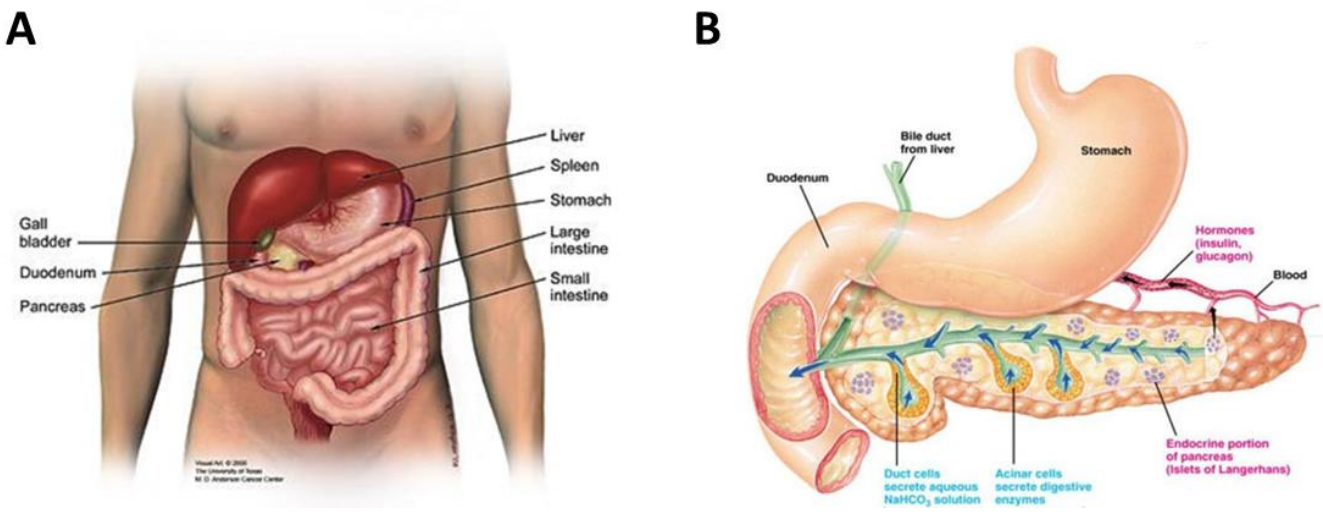


Figure 1.1 Location and characteristics of the human pancreas. A) The pancreas is located in the upper part of the abdominal cavity, behind the stomach. Figure adapted from: https://www.pancan.org/facing-pancreatic-cancer/about-pancreatic-cancer/what-is-the-pancreas/ **B)** The endocrine pancreas constitutes 1-2 % of the organ and consists of cell clusters called Islet of Langerhans that secrete hormones to the bloodstream. The exocrine pancreas constitutes the major volume of the pancreas and consists of acinar cells arranged in acini and duct cells that excrete digestive enzymes and $NaHCO_3$ to the duodenum, respectively. Figure adapted from: https://beyondthedish.wordpress.com/tag/pancreatic-ducts/

1.1.1 The exocrine pancreas

The exocrine pancreas consists of acini and the ductal system (Fig. 1.1B). The Acini have an irregular shape and contain acinar cells. The acinar cells synthesize, store and secrete digestive enzymes (Pandol, 2011). The acinar cell has two plasma membrane domains, the basolateral membrane and the apical membrane (Fig.1.2B) (Leung and Ip, 2006). The basolateral membrane is located towards the basal lamina (Fig. 1.2A) and contains receptors for neurotransmitters and hormones which stimulate the secretion of digestive enzymes. The

apical membrane is faced towards the acinar lumen and is connected to a tiny duct (Leung and Ip, 2006). The basal and the apical regions of the cell have different roles; the basal part contains the nucleus and endoplasmic reticulum for protein synthesis while the apical part store digestive enzymes in zymogen granules (Fig. 1.2B) (Pandol, 2011). The digestive enzymes are secreted into the lumen of the acinus which is connected to a ductal system (Pandol, 2011). Acinar cells are connected via the so-called centroacinar cells to ductal cells, which secrete bicarbonate ions (HCO_3^-), Na^+ and K^+ into the ducts (Fig.1.2A) (Junqueira and Carneiro, 2003, VanPutte et al., 2014). The exocrine secretions of the pancreas, which includes the digestive enzymes produced in the acinar cells along with HCO^-_3, N^+ and K^+ produced in the ductal cells, makes up the pancreatic juice (VanPutte et al., 2014).

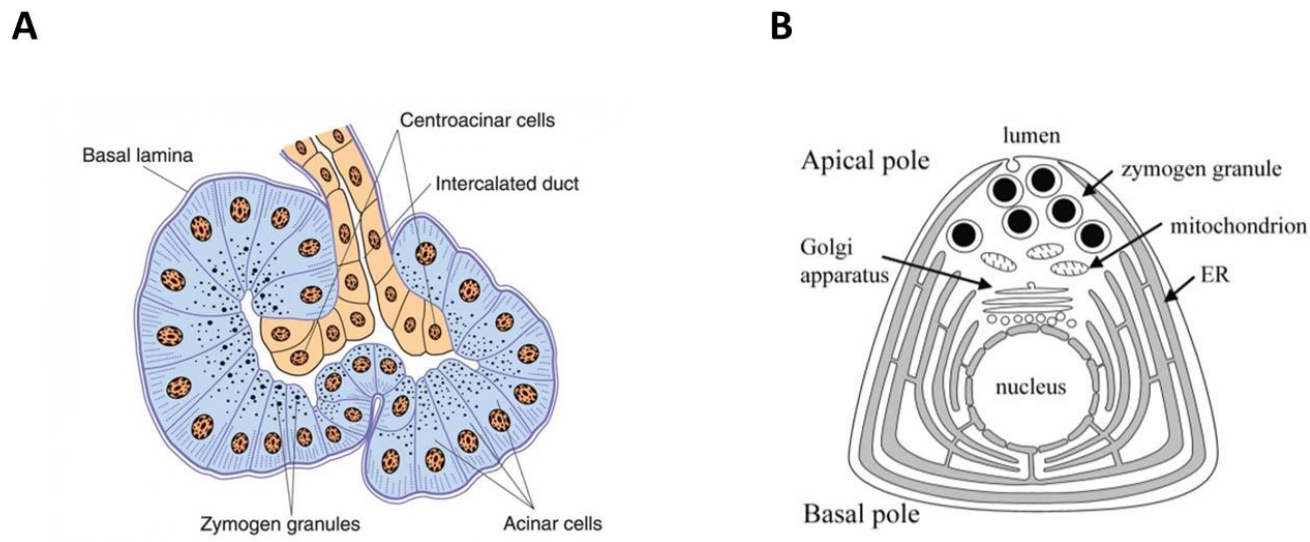


Figure 1.2 Structure of an acinus and an acinar cell. **A)** Schematic overview of the acinus in connection with the ductal system. Figure adapted from: https://www.diapedia.org/other-types-of-diabetes-mellitus/4104085136/acute-pancreatitis **B)** In the acinar cells, the digestive enzymes are stored in zymogen granules and released into the ducts when the pancreas is stimulated to secrete. The acinar cell has two plasma membrane domains, the apical- and basal membranes, respectively. Figure adapted from; (Burgoyne and Morgan, 2003).

The acinar cells secrete three main types of enzymes into the small intestine; proteases, amylases and lipases for the digestion of proteins, carbohydrates and fat, respectively (Holck, 2017). The proteases are generally made as inactive pro-enzymes to prevent self-digestion of the organ and are stored together with amylases and lipases in the zymogen granules of the acinar cells (Kumar et al., 2013). Deoxy-ribonucleases and ribonucleases degrade DNA and RNA to single nucleotides and are also components of the pancreatic juice (VanPutte et al., 2014).

Before the chyme (partially digested food) enters the duodenum, there is a slight increase in pancreatic secretion, due to an increased activity of the vagus nerve. The vagus nerve activates the secretion of gastrin, which is a hormone that is produced in the stomach. Further, gastrin stimulates the secretion of pepsin and hydrochloric acid (in the stomach). The major secretion of pancreatic juice, however, starts when the chyme enters the duodenum (Sand et al., 2011, Øyri, 2011). Two hormones that stimulate the exocrine pancreas are secreted from the duodenal epithelium: cholecystokinin and secretin. Cholecystokinin promotes secretion of the enzyme-rich fluid secreted by zymogen granules. Secretin stimulates the secretion of bicarbonate from the duct cells (Junqueira and Carneiro, 2003). The secretion of bicarbonate neutralizes the acidic stomach contents, giving the pancreatic enzymes an optimal pH for activity (Junqueira and Carneiro, 2003).

1.1.2 Diseases of the exocrine pancreas

Acute pancreatitis

Acute pancreatitis is an acute inflammatory disorder of the pancreas (Kumar et al., 2013). In the start of the disease, swelling and fluid retention in the organ is observed, which can further develop into necrosis and bleeding (Aabakken, 2016). During disease progression there is insufficient secretion of pancreatic juice to the intestine, which may result in a high volume of digestive enzymes in the pancreas. Consequently, this could lead to pre-activation of digestive enzymes in the pancreas and thus lead to self-digestion and inflammation (Aabakken, 2016).The main clinical feature of acute pancreatitis is abdominal pain, which can vary from mild to severe. The diagnosis is determined by exclusion of other causes of abdominal pain and elevated levels of lipase and amylase in blood plasma. The majority of acute pancreatitis cases are mild and reversible. However, 20 % of the cases of acute pancreatitis develop to severe diseases and may even be fatal (Kumar et al., 2013).

Chronic pancreatitis

Chronic pancreatitis is an long-standing inflammatory disease that results in an irreversible loss of pancreatic functions (Kumar et al., 2013, Kleeff et al., 2017). Long-term alcohol abuse and smoking are the main causes, but genetics also play a role in the pathogenesis (Kumar et al., 2013, Kleeff et al., 2017). The reason why alcohol causes chronic pancreatitis is not fully understood, but increased oxidative stress and direct injury to the acinar cells by inhibition of the endoplasmic reticulum (ER) activity may play a role (Kleeff et al., 2017). Chronic pancreatitis results in pain, progressive scarring of the pancreatic tissue and endocrine and

exocrine insufficiency due to loss of function of the parenchyma. The disease is associated with increased risk of developing pancreatic cancer. Recurrent attacks of acute pancreatitis may be the first stage in developing the disease. In the later stages, symptoms such as pain, steatorrhea and diabetes can be observed (Kleeff et al., 2017).

Hereditary pancreatitis

A rare cause of chronic pancreatitis is hereditary pancreatitis. This version of the disease is mainly due to germline mutations in the protease serine 1 gene (*PRSS1*). The penetrance of the disease is 80% and the inheritance pattern is autosomal dominant (Rebours et al., 2012). *PRSS1* encodes the enzyme cationic trypsinogen, expressed in the acinar cells of the pancreas. Trypsinogen is a precursor of trypsin 1. *PRSS1* mutations reduce the enzymatic activity or increase autocatalytic conversion of inactive trypsinogen to active trypsin 1 (Kleeff et al., 2017). Reduced inactivation of trypsin 1 and premature conversion of trypsinogen to trypsin are believed to cause pancreatic injury through autodigestion of pancreatic tissue, leading to an inflammatory response and recurrent acute pancreatitis which then progresses towards chronic pancreatitis (Kleeff et al., 2017).

Pancreatic cancer

The most common type of pancreas cancer is the ductal adenocarcinoma (PDAC), i.e. a malignant tumor with ductal differentiation originating from glandular tissue (Bertelsen, 2014, Hofsli, 2018). The first symptoms are vague, which often leads to a delayed diagnosis (Kumar et al., 2013). In the United States, pancreatic cancer has a 5 year survival of approximately 8% and is the third most common cause of cancer death (Tamura et al., 2018). The cause of the disease is not clearly understood, but pancreatitis, smoking and diet can contribute (Bertelsen, 2014). Inherited and acquired mutations in cancer-associated genes can also be a cause of pancreas cancer (Kumar et al., 2013).

In pancreatic ductal adenocarcinoma, DNA abnormalities are most frequently found in one oncogene and three tumor suppressor genes. Ninety % of the tumors contain somatic mutations in the KRAS proto-oncogene (*KRAS*). The most frequently altered tumor suppressor gene is cyclin dependent kinase inhibitor 2A (*CDKN2A*), which encodes an essential cell cycle regulator. This gene has a loss of function in more than 90 % of ductal adenocarcinomas. Tumor protein 53 (*TP53*) is also a tumor suppressor gene, and somatic mutations in this gene are common in pancreatic cancer. *TP53* encodes for a protein that is active in cellular stress response. Finally, SMAD family member 4 (*SMAD4*) is a tumor

suppressor gene that mediates signaling downstream of the transforming growth factor receptor (TGFβ). *SMAD4* is inactivated in approximately 50 % of tumors (Kamisawa et al., 2016). The cause of the molecular changes in the genes mentioned above is not known (Kumar et al., 2013).

Adenocarcinoma of the pancreas mostly affects the elderly, and the majority of the patients are between the ages of 60 to 80 (Kumar et al., 2013, Bertelsen, 2014). When a tumor is located in the pancreatic head region, there might be an obstruction of the distal common bile duct. In 50 % of pancreatic cancer cases there is a distention of the biliary tract, which may lead to jaundice (Kumar et al., 2013).Weight loss and abdominal pain are other symptoms seen in the late stages of the disease (Bertelsen, 2014).

1.1.3 The endocrine pancreas

The endocrine pancreas represents 1-2 % of the volume of the organ and consists of four major cell types; alpha-, beta-, delta- and pancreatic polypeptide (PP) cells (Fig.1.3). These cells are arranged in clusters called the Islets of Langerhans (Kumar et al., 2013). The mean diameter of an adult human islet is 140 μm (In'T Veld and Marichal, 2010). The different cell types of the endocrine pancreas all produce and secrete hormones that regulate the glucose level in the blood stream.

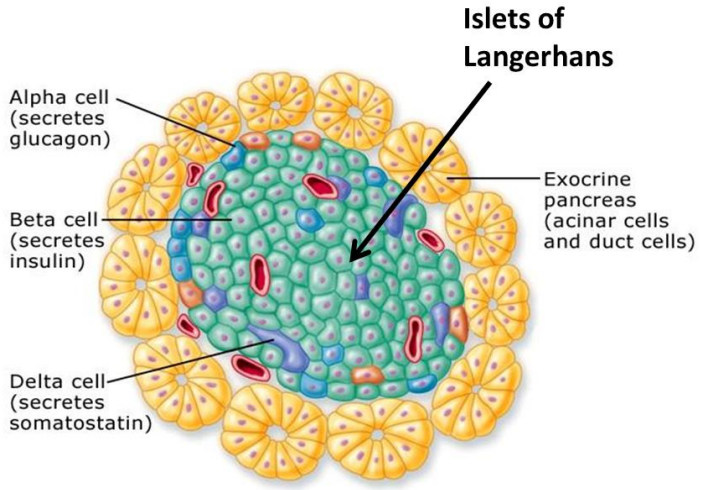


Figure 1.3 Features of the endocrine pancreas. The endocrine pancreas consists of four major cell types; beta, alpha, delta and PP cells. These cells are arranged in clusters called Islets of Langerhans. (Figure adapted from: https://www.wonderwhizkids.com/images/content/biology/human_physiology/endocrine_reproductive_system/conceptmap/Pancreas.html).

The beta cells produce insulin, while the alpha cells produce glucagon. A high level of glucose in the bloodstream promotes insulin secretion from the beta cells. The secretion of insulin will increase the uptake of glucose in the tissues and promote the production of glucose to glycogen in the liver. In contrast, a low level of glucose in the bloodstream will initiate the breakdown of glycogen, and thus increase the glucose level in the bloodstream (Holck, 2017).

The release of both insulin and glucagon are suppressed by the hormone somatostatin that is produced in the delta cells (Kumar et al., 2013). The PP cells contain a pancreatic polypeptide called vasoactive intestinal peptide (VIP). VIP have several gastrointestinal effects, such as inhibition of intestinal motility and secretion and stimulation of gastric and intestinal enzymes (Kumar et al., 2013).

1.1.4 Diseases of the endocrine pancreas

Diabetes mellitus

Diabetes mellitus is a common term for metabolic diseases that are characterized by chronic hyperglycemia (Kumar et al., 2013, Kharroubi and Darwish, 2015). Hyperglycemia is a result of insufficient insulin action, secretion, or both. The symptoms of the disease vary due to the duration and type of diabetes (Kharroubi and Darwish, 2015). The disease can be divided into two major groups; type 1 diabetes (T1D) and type 2 diabetes (T2D) (Kumar et al., 2013).

T1D was formerly called “juvenile onset diabetes” or “insulin-dependent diabetes”. This form represents 5-10 % of diabetes cases and is most common in children and adolescents. The disease is caused by destruction of the pancreatic β-cells due to an autoimmune response, which results in an absence of insulin secretion (Kharroubi and Darwish, 2015, American Diabetes Association, 2017). A hallmark of T1D is the presence of one or more autoimmune markers like autoantibodies towards glutamate decarboxylase (GAD), the tyrosine phosphatases IA-1, IA-2B and ZnT8, and insulin. Both environmental factors and genetics have a central role in the development of the disease. The treatment for patients with type 1 diabetes is insulin injections (Kharroubi and Darwish, 2015).

T2D represents 90-95 % of all diabetes cases. The disease was previously called “adult onset diabetes” or “noninsulin dependent diabetes”. Individuals with T2D exhibit a peripheral insulin resistance and relative insulin deficiency (American Diabetes Association, 2017). This form of the disease does not include autoimmune β-cell destruction. The risk for developing

the T2D increases with body weight, age and lack of physical activity, and patients diagnosed with the disease are often obese. The patients develop hyperglycemia gradually but the symptoms may be silent enough for the disease to be undiagnosed for years. Weight reduction and/or pharmacological treatment may improve insulin resistance (American Diabetes Association, 2017). In the final phase of the disease, insulin injections may be necessary (Kharroubi and Darwish, 2015).

Monogenic diabetes

In addition to T1D and T2D, there are two other main categories of diabetes; gestational diabetes and "other specific types of diabetes" (American Diabetes Association, 2017). The latter includes monogenic diabetes, a group of diseases that affects the normal β-cell physiology or pancreas development (Molven and Njølstad, 2011). The disease is caused by a defect in a single gene, and mutations in around 20 genes are known to cause monogenic diabetes (Molven and Njølstad, 2011, Søvik et al., 2013). Neonatal diabetes and maturity-onset diabetes of the young (MODY) are the two main types of monogenic diabetes (Søvik et al., 2013).

1.2 Carboxyl ester lipase (CEL)

Carboxyl ester lipase (CEL) is one of four major lipases that are secreted by the exocrine pancreas (Johansson et al., 2018). The digestive enzyme is also referred to as bile salt-stimulated lipase (BSSL) (Hernell and Olivecrona, 1974) or bile salt-dependent lipase (BSDL) (Abouakil and Lombardo, 1989).

1.2.1 The CEL protein

CEL is mainly expressed in the acinar cells of the pancreas. The enzyme activity is stimulated by bile salts (hence its names BSSL/BSDL) and hydrolyzes cholesterol esters, dietary fat and fat-soluble vitamins (Lombardo et al., 1978, Lombardo and Guy, 1980, Bläckberg and Hernell, 1983, Hui and Howles, 2002). Of the total amount of detected proteins in pancreatic juice, CEL represents 4 % (Lombardo et al., 1978). CEL is also expressed in lactating mammary glands and secreted with the mother`s milk (Bläckberg et al., 1981).

The CEL protein contains two major structural domains: a globular N-terminal domain with a signal peptide and a catalytic site, and a C-terminal including a variable number of tandem repeat (VNTR) domain (Fig. 1.4) (Reue et al., 1991, Terzyan et al., 2000). Each VNTR repeat

consists of 11 amino acids (AA) and the number of repeats vary from 3-23, with 16 as the most common. In addition, the VNTR is enriched in the amino acids proline (P), glutamine (E), serine (S), and threonine (T) (PEST sequences). The C-terminal is tailed by the unique sequence KEAQMPAVIRF (Johansson et al., 2018). Altogether, CEL give rise to a protein of 745 AA with a predicted molecular mass of 79 kDa (Rogers et al., 1986, Johansson et al., 2018).

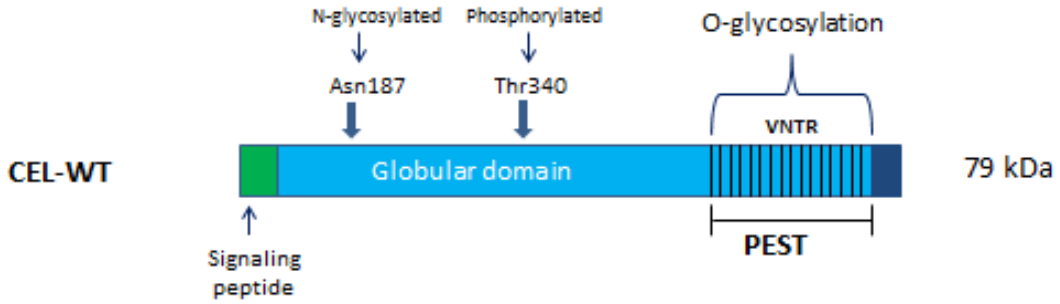


Figure 1.4 Structure of the carboxyl ester lipase protein. CEL has a theoretical size of 79 kDa. The protein contains two major structural domains; a globular N-terminal with a catalytic site and a signal peptide, and a C-terminal with a VNTR region (here illustrated with 16 repeats). CEL is N-glycosylated at Asn187 and phosphorylated at Thr340. The VNTR region contains O-glycosylation sites which masks the PEST sequence.

During the translation process of CEL, the polypeptide chain is released into the lumen of endoplasmic reticulum (ER) where it forms a folding complex with several chaperons after the cleavage of the N-terminal hydrophobic signal peptide (Lombardo, 2001, Johansson et al., 2018). The protein is N-glycosylated in the ER, at residue Asn187 (Fig. 1.4). This modification is important for secretion and correct folding of CEL (Abouakil et al., 1993). The polypeptide is thereby transported to the Golgi apparatus, where the C-terminus is heavily O-glycosylated at serine and threonine residues in the PEST sequence (Bruneau et al., 1997). It has been reported that PEST can be a signal for rapid protein degradation (Rogers et al., 1986). Thus, O-glycosylation of CEL could play a protective role by masking these PEST sequences (Loomes et al., 1999).

CEL is phosphorylated at residue Thr340, after it has been completely glycosylated in the Golgi apparatus (Fig. 1.4). Phosphorylation of Thr340 allows final translocation through the secretory pathway (Pasqualini et al., 2000). The protein is then stored with other digestive enzymes in zymogen granules (Johansson et al., 2018). Some CEL molecules remain in complex with the molecular chaperone glucose-regulated protein with a size of 94 kDa (GRP94) when secreted into the duodenum (Bruneau et al., 1998). By renal filtration, the

protein is cleared from circulation, and can be detected in the urine of healthy individuals (Comte et al., 2006).

1.2.2 The *CEL* gene

The human *CEL* gene consists of 11 exons, covers approximately 10 kilobase pairs (kbp) of genomic sequence and is located on chromosome band 9q34.3 (Fig. 1.5) (Taylor et al., 1991, Lidberg et al., 1992). The VNTR is found in the last exon of the *CEL* gene, and this region consists of almost identical 33 base-pair (bp) segments each encoding the 11 AA repeats (Johansson et al., 2018). Located 11 kbp downstream and in tandem with *CEL* is the *CEL*-like pseudogene *CELP* (Lidberg et al., 1992, Madeyski et al., 1998). *CELP* lacks exons 2-7 of CEL, and is not expected to translate into a functional protein (Nilsson et al., 1993). The promotor region of the mouse *CEL* gene is more similar to the promotor region of the *CEL* pseudogene than to *CEL*. Thus, it have been suggested that *CELP* is the original gene (Madeyski et al., 1998).

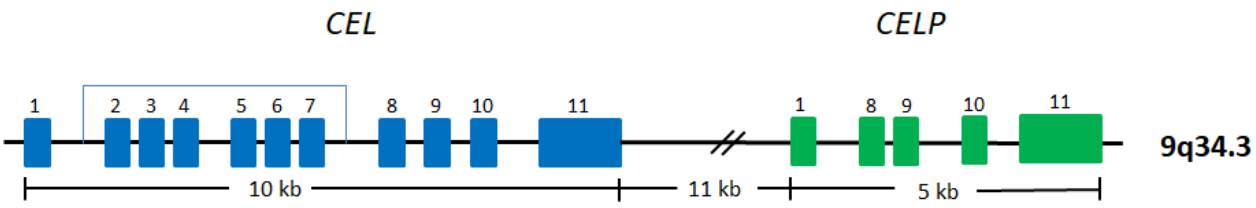


Figure 1.5 Structure of the human carboxyl ester lipase locus. The human *CEL* locus is located on chromosome 9q34.3. The *CEL* gene (blue) consists of 11 exons and covers 10 kbp. *CELP* is a *CEL*-like pseudogene that that lacks exons 2-7 and is located 11 kbp downstream and in tandem, with CEL.

1.2.3 Variants of the *CEL* gene

Due to the VNTR region, the *CEL* gene is highly polymorphic (Torsvik et al., 2010, Johansson et al., 2018). In addition to VNTR-length variation, single nucleotide polymorphisms (SNPs), single base insertions and single base deletions have been reported. Furthermore, copy number variants (CNVs) of the *CEL* locus have been described (Johansson et al., 2018).

VNTR length variants

As mentioned above, the number of repeats in the VNTR region in human *CEL* vary between 3-23, but 16 repeats are the most common (Johansson et al., 2018). The repeats differs also between species with examples like: no repeats in chicken, 3-4 repeats in rodents, 13 in dog and 39 repeats in gorilla (Holmes and Cox, 2011).

SNP

The SNP denoted rs488087 causes a change from C to T in the second VNTR repeat. This SNP have been proposed to result in an increased risk of developing pancreatic cancer (Martinez et al., 2015). However, only 30 patients with PDAC were included in this study. Thus, to conclude that this SNP gives an increased risk for PDAC, a significantly larger patient cohort has to be analyzed.

INS

Single base insertions in the VNTR region result in premature stop-codons and truncated CEL proteins. These mutations have been associated with fecal elastase deficiency in diabetic patients (Ræder et al., 2006). Insertions in repeat 4, 7, 8, 9, 10 and 11 have been observed (Johansson et al., 2018).The combined allele frequency of single base insertions is ca. 0.07 in normal controls of Northern European descent. Thus, the elimination of the C-terminal amino acid sequence KEAQMPAVIRF do not seem to have a high biological impact (Ræder et al., 2006).

The insertion of an extra cytosine in the CEL VNTR will give rise to a new C-terminal amino-acid sequence; PRAAHG (normally: PAVIRF) (Fig. 1.6). Such insertions have been studied by Martinez et al. in pancreatic tissue specimens and in HEK293 cells (Martinez et al., 2016). They developed an antibody directed towards the new C-terminal PRAAHG sequence to be able to detect CEL insertion variants. Further, they suggested that CEL insertions may arise as early somatic mutations in PDAC and that the detection of the modified C-terminal could be utilized as a diagnostic tool (Martinez et al., 2016).

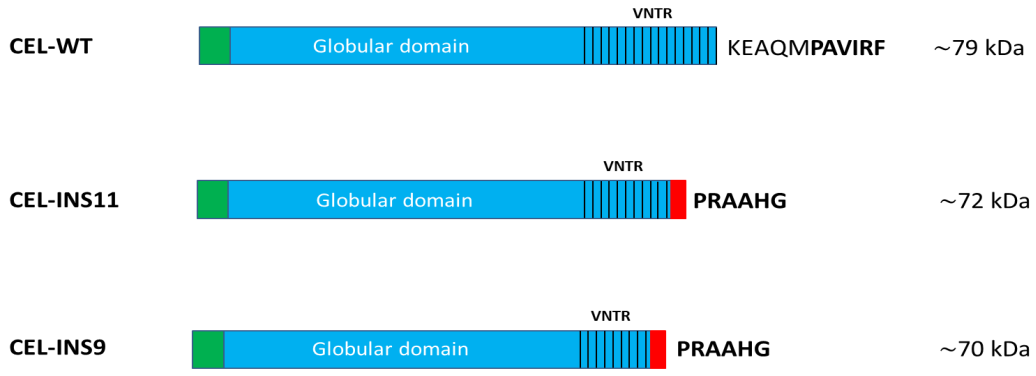


Figure 1.6 Structure of CEL insertion variants. Insertion of a cytosine in the VNTR will lead to a premature stop-codon and a truncated protein, and the C-terminal will end with the six amino acids, PRAAHG. In this figure, the insertion variants CEL-INS11 and CEL-INS9 are represented together with CEL-WT. The 11 last amino acids of the CEL-WT are shown.

DEL

Single base deletions in the VNTR result in a frameshift, premature stop codons and a truncated C-terminal. Two deletion variants of *CEL* have been observed with a deletion in the first (DEL1) and fourth (DEL4) repeat, respectively. DEL1 is predicted to change 110 amino acids in the C-terminal, while DEL4 is predicted to change 99 amino acids in the C-terminal (Johansson et al., 2018). CEL-DEL1 and CEL-DEL4 are causative mutations for the CEL-MODY syndrome (Ræder et al., 2006) that is described in more detail in section 1.3.

CNVs

The probability of genomic rearrangements is increased with the constellation of a gene close to a homologous sequence. Thus, it can result in formation of a fusion gene that has different functional properties than the original. These fusion genes may be associated with disease (Harel and Lupski, 2017).

CNVs of *CEL* have been reported, both a duplication and a deletion hybrid allele, which apparently are a result from nonallelic homologous recombination between *CEL* and *CELP*. The *CEL* duplication allele is not expected to translate into a protein due to a premature stop codon in *CELP* exon 8. However, the deletion hybrid allele (*CEL-HYB*) includes three VNTR repeats from the *CEL* pseudogene and encodes a CEL-CELP fusion protein that has been associated with chronic pancreatitis (Fig.1.7) (Fjeld et al., 2015).

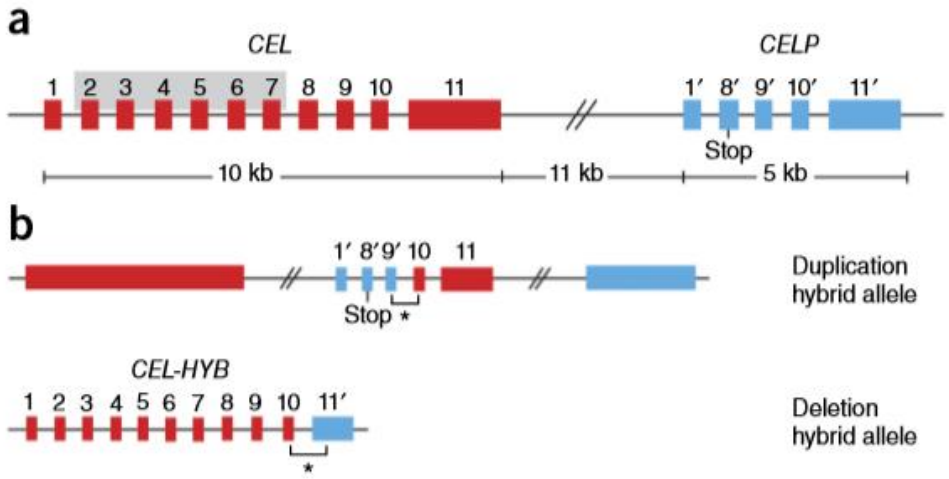


Figure 1.7 Copy number variants of the human *CEL* gene: duplication and a deletion variant. a) Structure of the human CEL locus consisting of *CEL* and the pseudogene *CELP*. The gray shading of exon 2-7 in *CEL* corresponds to the exons missing in *CELP*. b) Structure of a duplication allele and a deletion hybrid allele (*CEL-HYB*) of *CEL*. The duplication allele consist of exon 1, 8 and 9 from *CELP* and exon 10-11 from *CEL*. The deletion hybrid allele consists of exon 1-10 of *CEL* and exon 11 of *CELP*. Figure adapted from (Fjeld et al., 2015).

1.3 Carboxyl ester lipase in human disease

1.3.1 MODY8

MODY8 is a type of monogenetic diabetes caused by a single base-pair deletion (DEL1 or DEL4) within the *CEL*-VNTR. The disease causes pancreatic exocrine dysfunction, lipomatosis and pancreatic cysts in addition to diabetes. The disease is autosomal dominant inherited (Johansson et al., 2018).

The MODY8 syndrome was first discovered in a Norwegian family suspected to have monogenic diabetes (Ræder et al., 2006). The family members that were affected fulfilled the criteria for MODY, which are primary beta cell dysfunction and autosomal dominant diabetes detected before 40 years of age. However, the individuals also experienced, relatively early in life, fecal elastase deficiency due to pancreatic exocrine dysfunction, a high content of fat in the pancreas and repeated attacks of abdominal pain. By the age of 40, most of them had also developed diabetes and pancreatic cysts (Ræder et al., 2006, Ræder et al., 2014). The CEL-DEL1 and DEL4 mutation results in an altered reading frame and an aberrant C-terminus of the CEL protein (Ræder et al., 2006). It has been suggested that MODY8 might be a protein misfolding disease due to the formation of both intra- and extra-cellular protein aggregates in cell model systems (Johansson et al., 2011, Molven et al., 2016).

1.3.2 Chronic pancreatitis

CEL-HYB is a fusion gene between *CEL* and *CELP* (Fig. 1.7) and is identified as a genetic risk factor for chronic pancreatitis. In patients from Germany and France with idiopathic chronic pancreatitis, the carrier frequency of *CEL-HYB* was increased by >5 fold, compared to the general population. However, even though the allele is a significant risk factor for developing chronic pancreatitis, the majority of the *CEL-HYB* carriers in the general population (about 1%) are expected to stay healthy (Fjeld et al., 2015, Molven et al., 2016). An association between longer *CEL* VNTR repeats and alcoholic chronic pancreatitis were reported by a Japanese study (Miyasaka et al., 2005), but the results could not be replicated in two larger studies from Europe (Ragvin et al., 2013, Fjeld et al., 2016). The varying results between European and Asian studies could be due to methodical issues or ethnic differences (Johansson et al., 2018).

1.3.3 Other pancreatic diseases

The *CEL* gene has also been investigated in the context of pancreas cancer and type 1 diabetes. One study looked at a possible connection between VNTR length and pancreas cancer, but no association was found (Dalva et al., 2017). Further, as mentioned above, associations between pancreas cancer and insertion variants in the *CEL* VNTR (Martinez et al., 2016) or the SNP (rs488087) in *CEL* exon 11 (Martinez et al., 2015) have been suggested. Insertion variants in the VNTR region have also been associated with exocrine dysfunction in type 1 diabetes patients (Ræder et al., 2006).

1.3.4 Non-pancreatic diseases

It have been shown in an HIV infected patient group that has a low/high VNTR repeat genotype is associated with accelerated disease progression, while a high/high VNTR repeat genotype is related to a slower disease progression (Stax et al., 2012).

Another study compared serum lipid phenotype and CEL genotyping and revealed a correlation between the serum cholesterol profile and the number of VNTR repeats. Low-density lipoprotein (LDL) cholesterol levels in individuals that carried at least one allele with less than 16 repeats, which are the most common, were significantly lower compared to individuals carrying two 16 repeats. The authors suggested as a possible explanation that CEL proteins with less proline-rich repeats are less well protected from degradation in the intestinal tract and/or secreted at a lower rate. Thus, the hydrolysis and uptake of cholesterol dietary lipids are less efficiently assisted (Bengtsson-Ellmark et al., 2004).

2. Aims of the study

The overall aim for this study was to understand the biological impact of insertion variants in the *CEL* gene.

The specific objectives of the project were:

1. To analyze the location and frequency of *CEL* insertion variants in human samples from a Norwegian pancreas biobank
2. To establish a strategy for detecting CEL insertion variants both at the DNA and protein level
3. To examine the impact of CEL insertion variants

3. Materials

Table 3.1 Long-Range PCR and agarose gel electrophoresis

Material	Catalog Number	Supplier
E.Z.N.A ® Tissue DNA kit	D3396	OMEGA Bio-Tek
NuSieve GTG (agarose)	50084	LONZA
TBE Buffer x10	A3945, 1000	PanReac, AppliChem
2XGC-Buffer 1	SD1432	TaKaRa
dNTP-mix (2.5 mM each)	SD0316	TaKaRa
Betain solution (5M)	B0300-1VL	Sigma-Aldrich
LaTaq (5U/µl)	#RR02AG	TaKaRa
Ethidium Bromide (0.625 µg/µl)	E406-15ml	VWR
Gel Loading Buffer	G2526	Sigma-Aldrich
1 Kb marker	#N3232S	New England Biolabs

Table 3.2 Sanger sequencing

Material	Catalog Number	Supplier
Sephadex ® G-50 Superfine	G5050-50G	Sigma-Aldrich
Multiscreen-HV Filter plate	MAHVN4510	Millipore
BigDye Xterminator™ Purification kit	4376486	Thermo Fisher Scientific
BigDye Terminator v.3.1 Cycle sequencing kit	4336917	Thermo Fischer Scientific
Big Dye terminator v.1.1, v.3.1 sequencing buffer	4336697	Thermo Fischer scientific
Big Dye terminator v.1.1 Cycle sequencing kit	4336774	Thermo Fischer Scientific
Illustra ™ExoProStar	US77705V	Sigma-Aldrich
MicroAmp™ Optical 96-well reaction plate	N8010560	Thermo Fischer scientific

Table 3.3 Plasmids

Plasmids	Variant	Description
pcDNA 3	CEL-WT[1]	Plasmid expressing CEL-WT with 16 repeats of the VNTR
pcDNA 3	CEL-INS9[1]	Plasmid expressing CEL with insertion in repeat 9 of the VNTR
pcDNA 3	CEL-INS11[1]	Plasmid expressing CEL with insertion in repeat 11 of the VNTR
pcDNA 3	CEL-TRUNC[1]	Plasmid expressing CEL lacking the VNTR region
pCDNA 3	Empty vector	Invitrogen

[1]The plasmids CEL-WT, CEL-INS9, CEL-INS11 and CEL-TRUNC were provided as a gift from Prof. M. Lowe, Washington University School of Medicine, St. Louis USA.

Table 3.4 Plasmid preparation and purification

Material	Catalog Number	Supplier
imMedia™ Growth Medium, agar, ampicillin	Q60120	Thermo Fischer Scientific
Lysogeny broth (LB) medium	L7275-500TAB	Sigma-Aldrich
Super optimal broth with catabolite repression (SOC) medium	15544-034	Thermo Fischer Scientific
Ampicillin sodium salt	A9518-5G	Sigma-Aldrich
Plasmid Midi kit (100)	12245	Qiagen
E.coli OneShot® TOP10 chemically competent cells	C404010	Thermo Fischer Scientific

Table 3.5 Cell culture

Material	Catalog Number	Supplier
Dulbecco`s Modified Eagle Medium (DMEM), high glucose, pyruvate	41966-029	Thermo Fischer Scientific
Fetal Bovine Serum (FBS)	F7524-500ml	Sigma-Aldrich
Dulbecco`s Phosphate buffered saline (PBS)	D8537-500ml	Sigma-Aldrich
Antibiotic Antimycotic	15240062	Thermo Fischer Scientific
Trypsin EDTA 1x (0.05%)	25300-54	Thermo Fischer Scientific

Table 3.6 Transient transfection

Material	Catalog Number	Supplier
Lipofectamine™ 2000 Transfection reagent	11668-019	Thermo Fischer Scientific
OPTI-MEM (1x) Reduced Serum Medium	31985-062-100 ml	Thermo Fischer Scientific
Complete™ Mini EDTA-free protease inhibitors cocktail tablets	11836170001	Sigma-Aldrich
RIPA lysis buffer (10x)	20-188	Millipore
Falcon™ Polystyrene Microplates (6-well plate)	08-772-29	Fischer Scientific

Table 3.7 Immunocytochemistry

Material	Catalog Number	Supplier
Falcon™ Polystyrene Microplates (12-well plate)	08-772-29	Fisher Scientific
Poly-L-Lysine 0.01% solution	P4832-50ml	Sigma-Aldrich
Triton x100	437002A	VWR
Slides, microscope	S8902	Sigma-Aldrich
Superfrost microscope slides	Z692255-100EA	Sigma-Aldrich
Normal Goat Serum Control	10000C	Thermo Fischer Scientific
Paraformaldehyde (PFA)	818715	Millipore

3.8 Immunohistochemistry

Material	Producer	Catalogue number
DAKO pen	S2002	Agilent DAKO
EnVision+ Kits, HRP. Mouse (DAB+)	K400711-2	Agilent DAKO
EnVision+ Kits, HRP. Rabbit (DAB+)	K40111-2	Agilent DAKO
Envision Flex Hematoxylin	K8008	Agilent DAKO

Table 3.9 SDS-PAGE and Western blot

Material	Catalog Number	Supplier
Pierce™BCA Protein Assay kit	23225	Thermo Fischer Scientific
NuPAGE ® LDS sample buffer (4x)	NP0007	Thermo Fischer Scientific
NuPage ® Sample Reducing agent (10x)	NP0009	Thermo Fischer Scientific
NuPage ® MOPS SDS Running buffer (20x)	NP0001	Thermo Fischer Scientifc
NuPage ® Novex 10 % Bis-Tris protein gels 1.0 mm, 10 well	NP0301BOX	Thermo Fischer Scientific
Precision Plus Protein™ dual color standard	161-0374	BioRad
Magic Mark XP western protein standard	LC5603	Thermo Fischer Scientific
Immobilion-P membrane, PVDF, 0.45µm	IPVH07850	Millipore
NuPAGE ® Transfer Buffer (20x)	NP0006-1	Thermo Fischer Scientific
Blocking grade blocker nonfat-dry milk	170-6404	BioRad
Phosphate buffer saline (PBS) tablets	18912-014	Thermo Fischer Scientific
Tween®20	8.22184.0500	Millipore
Restore™Western Blot Stripping Buffer	21059	Thermo Fischer Scientific
Pierce™ ECL Plus Western Blotting Substrate	32132	Thermo Fischer Scientific

Table 3.10 Buffers and solutions

Buffers and solutions	Use	Composition
1xTBE buffer	Gel electrophoresis	For 1 L: 100 ml TBE buffer 10x from PanReac dissolved in 900 ml MilliQ water
1% Agarose	Gel electrophoresis	1 gram NuSieve GTG (agarose) dissolved in 100ml 1xTBE buffer with 0.625 µg/µl EtBr
1 % Agarose gel	Gel electrophoresis	50 ml 1 % agarose with 0.625 µg/µl EtBr
NuPAGE® Transfer Buffer (1x)	Western blot	For 1 L: 50 ml NuPAGE ® Transfer buffer (20x) in 850 MilliQ dH_2O and 10 % methanol
NuPage ® SDS MOPS buffer (1x)	SDS-PAGE	For 1 L: 50 ml NuPAGE® MOPS buffer (20x) in 950 MilliQ dH_2O
PBS-Tween 0.05 %	Western Blot	1 PBS tablet dissolved in 500ml MilliQ dH_2O with 0.05 % Tween®20
5% milk	Western Blot	5 gram blocking grade blocker nonfat-dry milk and 100ml 0.05 % PBS-T
1x Ripa Lysis Buffer	Cell Lysis	1 ml 10xRIPA lysis buffer in 9 ml dH_2O + 1 tablet Complete^{TM} Mini EDTA-free protease inhibitors
Blocking buffer	Immunofluorescence	5% Normal goat serum in washing buffer
Washing Buffer (PBS 0.5 %)	Immunofluorescence	For 1L: 2 PBS tablets and 1 ml Tween®20 and 1000 ml MilliQ dH_2O
Permeabilization solution	Immunofluorescence	15 µl Triton X-100 and 15 µl Tween 20 dissolved in 14,97ml 1xPBS
Paraformaldehyde (PFA) solution	Immunofluorescence	6% PFA solution, 3 gram PFA dissolved in 50 ml MilliQ water + 5 drops of 1 M NaOH
0.2 M Phosphate Buffer (PB)	Immunofluorescence	0.2 M Na_2HPO_4 and 0.2 M NaH_2PO_4 adjusted to pH 7.2
Fixative	Immunofluorescence	6% PFA + 0.2 M PB pH 7.2 (1:1)
1mM EDTA – 10mM Tris buffer	Immunohistochemistry	1,205g Tris + 0,395g EDTA to 1L dH_2O, adjust pH to 9.0
Antibody diluent	Immunohistochemistry	0.05M Tris + 0.15M NaCl + 1 % BSA + 0.02% Na-Azid + 0.0 % Tween, pH adjusted to 7.4

Table 3.11 Antibodies

Antibody	Catalogue number	Supplier	Usage
Anti-CEL antibody (rabbit polyclonal)	-	Gift from Prof M. Lowe, Children's Hospital-Washington University in St. Louis USA	Primary (WB)*
As20 anti-CEL antibody (mouse monoclonal)	-	Gift from Prof O.Hernell, University of Umeå, Sweden	Primary (ICC)*
Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	A-21206	Thermo Fischer Scientific	Secondary (ICC)*
F(ab')2-Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	A-11017	Thermo Fischer Scientific	Secondary (ICC)*
GAPDH antibody (0411)	sc-47724	Santa Cruz Biotechnology	Primary (WB)*
Donkey anti-mouse IgG-HRP (polyclonal)	sc-2318	Santa Cruz Biotechnology	Secondary (WB)*
Goat anti-Rabbit IgG (H+L) HRP	#65-6120	Thermo Fischer Scientific	Secondary (WB)*
Anti-PAVIRF (rabbit polyclonal)	-	Gift from Profs. Dominique Lombardo & Eric Mas, Marseille, France	Primary (ICC, IHC and WB)*
Anti-PRAAHG (rabbit polyclonal)	-	Dominique Lombardo & Eric Mas, Marseille, France	Primary (ICC, IHC and WB)*
Anti-CEL (rabbit polyclonal)	HPA052701	Sigma-Aldrich	Primary (IHC)*
MACH 3 Rabbit HRP Polymer Medical	RH531H	Biocare Medical	Secondary (IHC)*
MACH 3 Rabbit Probe	RP531H	Biocare Medical	Secondary (IHC)*

*WB = Western blotting, ICC = Immunocytochemistry, IHC = immunohistochemistry

Table 3.12 Technical equipment

Instrument	Manufacturer
Applied Biosystems Thermal Cycler 2720	Thermo Fischer Scientific
GeneFlash Bio Imaging System	Syngene
3500xL genetic analyzer	Thermo Fischer Scientific
NanoDrop ND-1000	Thermo Fischer Scientific
Sceptre^{TM}2.0 handheld automated cell counter	Millipore
LAS-1000 Imager	Fujifilm
Leica Confocal SP5	Lecia Microsystems
Leica DM2000 LED	Leica Microsystems
Leica DMLB	Leica Microsystems

Table 3.13 Analytical software

Analytical software	Supplier
SeqScape v. 2.7	Applied Biosystems
Adobe Photoshop CS5 imaging	Adobe Photoshop
LASV4.8	Leica
ZEN 2011	Zeiss
Image Gauge v4.0	Fujifilm

4. Methods

4.1 Patient material

4.1.1 Pancreatic Biobank

The human samples used in this study were from the Pancreas Biobank at Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen. This biobank consist of blood and tissue samples from approximately 500 patients evaluated at the Department of Gastrointestinal Surgery, Haukeland University Hospital. The samples are collected on the basis of suspected or possible pancreatic tumor. The Research Ethics Committee (REK) of Western Norway has approved the study (no. 2013/1772) and the patients have given their informed consent. For 300 of the biobank samples, *CEL* VNTR length had been determined previously (Dalva et al., 2017), and the samples of this thesis were selected among those.

4.2 Genotyping of the *CEL* – VNTR region

4.2.1 Purification of human DNA samples

Purification of genomic DNA from blood samples provided by the Pancreas Biobank was performed by Chief Engineer Solrun Steine using E.Z.N.A ® Tissue DNA kit.

4.2.2 Polymerase Chain Reaction (PCR)

Long range PCR was performed to amplify *CEL* exon 8-11 of genomic DNA from samples in the Pancreas Biobank. The PCR master mix was prepared as described in Table 4.1, including 10 ng of DNA and *CEL*-specific primers. The primers are listed in Table 4.2 and the primer binding sites are illustrated in Figure 4.1 (for a more detailed illustration of the binding sites, see Appendix 1). The conditions for thermocycling are presented in Table 4.3, and were performed on an Applied Biosystems 2720 Thermal Cycler.

Table 4.1 PCR master mix

Component	Volume (µl)
2XGC-buffer 1	6.3125
dNTP-mix (2,5mM each)	2
L11F (20 µM)	0.375
VNTR-R (20 µM)	0.375
MilliQ dH_2O	1.375
Template (10 ng/µl)	1
Betain (5M)	2.5
LaTaq (5U/µl)	0.0625
Total	**14 µl**

Table 4.2 CEL primers for Long Range PCR and Sanger sequencing

Analysis	Primer	Sequence (5`→3`)
Long-range PCR	L11F	5`-GTGCCTCACTCATTCTTCTATGGCAAC-3`
	VNTR-R	5´- TCCTGCAGCTTAGCCTTG GG -3´
Sequencing PCR	EF	5`-CACACACTGGGAACCCT-3`
	VNTR-R	5`-TCCTGCAGCTTAGCCTTGGG -3`

Table 4.3 Thermal amplification cycles and conditions

Step	Temperature (°C)	Time	Number of cycles
Initial denaturing	94	1 min	
Denaturing 1	94	20 sec	14
Annealing 1	60	10 min	
Denaturing 2	94	20 sec	20
Annealing 2	62	10 min	
Extension	72	10 min	
	4	∞	

4.2.3 Agarose gel electrophoresis

The Long Range PCR products were verified by agarose gel electrophoresis. Samples were separated on a 1 % agarose gel containing 0.625 µg/µl ethidium bromide. For each sample, 6 µl of PCR product was mixed with 3.5 µl of 6x loading buffer before loading on the gel. In addition, 5 µl of 1 kb DNA marker was used as a molecular-weight size marker. The gel was run in 1xTBE buffer for ca. 1 hour and 15 minutes at 80 V. For visualization, the Gene Flash Bioimaging system was used.

4.2.4 Sanger sequencing of the *CEL* VNTR region

The long-range PCR products were treated with ExoSap before sequencing to remove excess of nucleotides. A total of 2 µl ExoSap were added to 5 µl of PCR product and incubated for 15 min at 37 °C and 15 min at 80 °C.

For Sanger sequencing, 2 µl of the ExoSap mix were used as template. The master mix for sequencing was prepared as described in Table 4.4. The primers are listed in Table 4.2 and the primer binding sites illustrated in Figure 4.1 (for a more detailed illustration of the binding sites, see Appendix 2). The conditions for thermocycling are presented in Table 4.5, and were performed on Applied Biosystems 2720 Thermal Cycler. The Sanger sequencing workflow is shown in Figure 4.2.

Table 4.4 Master mix for sequencing

Component	Volume (µl)
Primer (5 µM)	0.25
Betain (5M)	2
MilliQ dH_2O	2.75
Seq.Buffer	2
BD v.3.1	1
Templat	2
Total	**10 µl**

Table 4.5 Thermal amplification cycles and conditions for sequencing

Step	Temperature (°C)	Time	Number of cycles
Initial denaturing	96	10 min	
Denaturing 1	96	10 sec	25
Annealing 1	58	5 sec	
Extension	60	4 min	
	4	∞	

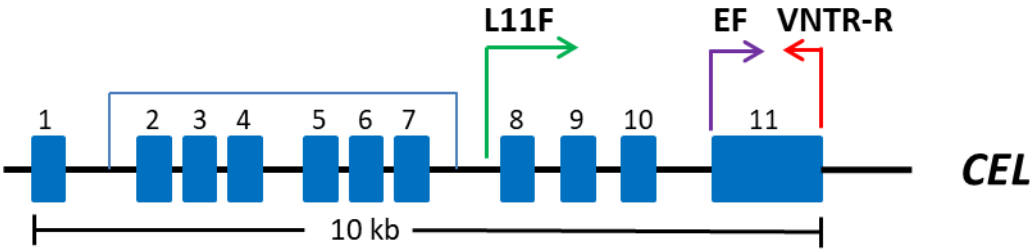


Figure 4.1 Primer binding sites. Primer binding sites for forward (L11F) and the reverse (VNTR-R) strand synthesis in the Long Range PCR reaction. L11F binds in intron 7, and VNTR-R binds in intron 11. L11F and VNTR-R amplify a region of approximately 3800 base pairs. Primer binding sites for forward (EF) and reverse (VNTR-R) for the two sequencing reactions performed. EF binds in exon 11 and VNTR-R binds in intron 11. EF and VNTR-R cover a region of approximately 800 base pairs.

The labelled sequencing products were purified by using Sephadex ® G-50 Superfine, which is a gel filtration medium. Sephadex ® G-50 Superfine was transferred to a multiscreen-HV filter plate together with 300 µl MilliQ water, and incubated at RT for two hours. Next, the excess of MilliQ water was removed from the gel filtration medium by centrifugation for 5 min at 910 x g. The sequencing products were added 15 µl of milliQ and transferred to the packed multiscreen plate, placed on a sequencing plate, and centrifuged for 5 min at 910 x g. The purified sequencing products were analyzed on a 3500xL Genetic Analyzer. The sequences were evaluated manually.

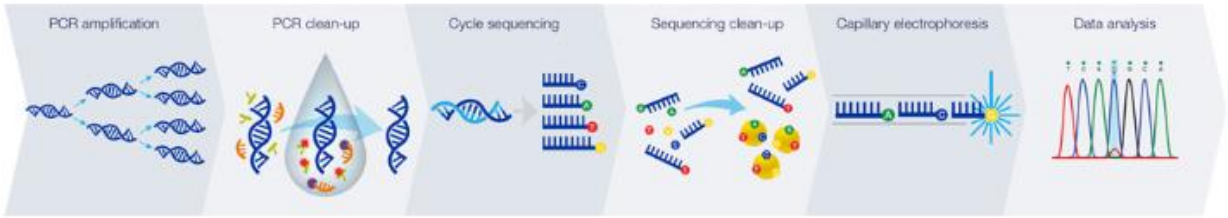


Figure 4.2 Sanger Sequencing workflow. The figure is adapted from Thermo Fischer scientific, from https://www.thermofisher.com/no/en/home/life-science/sequencing/sanger-sequencing/sanger-sequencing-workflow.html

4.3 Preparation and sequencing of CEL-expressing plasmids

4.3.1 Transformation of bacteria

CEL-expressing plasmids (Table 3.3) were transformed into OneShot ® TOP10 chemically competent *E.coli* cells by the heat-shock method. For each transformation, 1 µl of plasmid was added to 35 µl of *E.coli* cells, incubated on ice for 30 minutes and heated for 30 seconds at 42°C. After the heat-shock, the cells were incubated on ice for 2 minutes. Next, 250 µl of pre-warmed SOC medium (42°C) was added to the vial before incubation at 37°C and shaking at 250 rpm. After 1 hour, 100-200 µl of the *E.coli* cells was spread onto agar plates with ampicillin (100 µg/ml). The plates were incubated overnight at 37°C.

4.3.2 Bacterial cultures and plasmid purification using QIA filter Plasmid midi kit

After transformation, one bacteria colony was picked and inoculated in a starting culture containing 5 ml LB medium and 50 µl of ampicillin (100 µg/ml). The starting culture was incubated at 37 °C with shaking (250 rpm) for 8 hours. After 8 hours the starting culture was diluted 1:1000 in 25 ml with LB medium containing ampicillin (100 µg/ml). Then, the cultures were incubated for 16 hours at 37°C, shaking at 250 rpm. The next day, the cultures were transferred to a 50 ml falcon tube. Glycerol stocks were made for each culture. To make the glycerol stocks, 500 µl of bacterial culture were diluted with 500 µl of glycerol 50 %. The glycerol stocks were stored at -80 °C. The remaining cultures were centrifuged for 40 minutes at 4600 rpm. The supernatant was discarded, and the pellets were stored at -20°C until further use.

Plasmid purification was performed according to the Qiagen(R) Plasmid Midi Kit (100) protocol. Buffers are those provided in the kit. Briefly, the bacterial pellet was resuspended in 4 ml buffer P1 (added LyseBlue) and 4 ml buffer P2 was added. The mixture was incubated at room temperature (RT) for 5 minutes. Then, 4 ml of P3 buffer was added to the lysate. The lysate were poured in QIA filter cartridge and incubated at RT for 10 minutes. The cell lysate were filtered into a buffer-equilibrated QIAGEN-tip. After washing, DNA was eluted with 5 ml QF buffer. The eluted DNA was precipitated by adding 3.5 ml isopropanol and centrifuged at 4600 rpm for 45 minutes. After centrifugation, the DNA pellet was washed with 2 ml RT 70 % ethanol and centrifuged at 4600 rpm for 15 minutes. The supernatant was removed and

the DNA pellet was air-dried for 5-10 minutes. DNA was re-dissolved in 200 µl TE buffer and stored at RT overnight.

4.3.3 Determination of plasmid concentration and quality

To determine the concentration of the purified plasmids, the absorbance of 1 µl of DNA was measured at 260 nm on a NanoDrop®ND-1000 spectrophotometer. The plasmid quality was verified by agarose gel electrophoresis. A total of 2 µl (600 ng) DNA was separated on a 1 % agarose gel with 0.625 µg/µl ethidium bromide to visualize the DNA migration in the gel.

4.3.4 Sequencing of plasmid constructs

The plasmids expressing CEL-INS9 and CEL-INS11 included an insertion of a cytosine in the 9^{th} and 11^{th} VNTR repeat, respectively. To verify these insertions, the plasmid constructs were sequenced by Sanger sequencing. To the PCR reaction mix (Table 4.6), 600 ng DNA was added together with one of the *CEL* primers listed in Table 4.7. The thermocycling conditions are the same as described in Table 4.5

Table 4.6 Reaction mix for sequencing the CEL gene in plasmids

Component	Volume (µl)
Primer (5 µM)	0.25
Betain (5M)	2
MilliQ dH_2O	2.75
Seq.Buffer	2
BD v.1.1	1
Template	2
Total	**10µl**

Table 4.7 Primers used for sequencing the plasmids

Primer	Binding site	Sequence (5`→3`)
T7 (forward)	T7 promotor	ATTATGCTGAGTGATATCCC
BF (forward)	Exon 2	GAAGCTGGGCGCCGTGT
CF (forward)	Exon 4	TCACCTTCAACTACCGT
BR (reverse)	Exon 6	TGGCTCGCCGGATG
DF (forward)	Exon 8	CCGCCGACATCGACTA
CR (reverse)	Exon 8	CTCCTCCGTGACTTTC
DR (reverse)	Exon 11	GCCGCTGTTTTCCGTA
EF (forward)	Exon 11	CACACACTGGGAACCCT
BGH (reverse)	BGH polyadenylated sequence	ATCTTCCGTGTCAGCTCC

The plasmid sequencing products were purified by the Xterminator kit. A total of 45 µl of SAM solution buffer and 10 µl Xterminator mix was added to a sequencing plate. The sequencing products were added 15 µl of MilliQ water, and 10 µl of the volume was added to the sequencing plate. The plate was shaken for 20 minutes at 1800 rpm, and centrifuged for 2 minutes at 1000 x g. The samples were analyzed on a 3500xL genetic analyzer. The sequences were evaluated with SeqScape software v. 2.7.

4.4 Cell culturing and transfection

4.4.1 Culturing of human embryonic kidney cells 293 (HEK293)

For the cell culture experiments, HEK293 human embryonic kidney cells were employed (Clontech catalog no. 632180). The cells were cultured in Dulbeccos Modified Eagle`s medium (DMEM). DMEM was supplemented with 10 % fetal bovine serum and 100 U/ml of antibiotics antimycotic. The cells were grown at 37° with 5 % $CO_{2,}$ in a humidified atmosphere.

4.4.2 Thawing

HEK293 cells were thawed quickly in a 37°C water bath. The cells were transferred to a 15 ml tube with 12 ml of pre-warmed (room temperature) DMEM, and centrifuged at 1000 rpm for 5 minutes. The supernatant was removed and the pellet resuspended in 5 ml (T25cm^2) or 10 ml (T75cm^2) pre-warmed DMEM.

4.4.3 Sub-culturing and seeding

To passage the HEK293 cells, DMEM was removed and cells were washed in phosphate buffered saline (PBS). The PBS was removed and 1 ml trypsin EDTA (0.05%) was added to detach cells from the surface. The cells were resuspended in growth medium and further diluted in a suitable amount of DMEM before transferred to a T75 cm^2 flask. The flask was then filled up to 12 ml with fresh DMEM. PBS, DMEM and trypsin were all pre-warmed to RT before washing/passaging cells. For sub-culturing and seeding, only T75cm^2 flasks were used, unless stated otherwise.

4.4.4 Freezing

HEK293 cells were grown to 80-90 % confluency in a T75cm^2 flask prior to freezing. Complete growth medium was removed and cells were washed with 7 ml pre-warmed PBS. The PBS was removed, 3 ml Trypsin EDTA (0.05%) was added and the cells were resuspended in 5 ml DMEM. Next, the cells were transferred to a 15 ml tube and centrifuged for 5 minutes at 1000 rpm. The supernatant was removed and the pellet was resuspended in 3 ml freezing medium (1 ml of 100% DMSO and 9 ml complete media). The cells were divided into 3 vials containing 1 ml each. The freezing vials were transferred to a CoolCell™ LX Freezing Container and stored at -80°C o/n. For long-time storage, the vials were stored in a liquid nitrogen tank.

4.4.5 Transient transfection of HEK293 cells

Prior to transfection, cells were grown to 60-70 % confluency in a T75cm^2 flask. The cells were washed in PBS, trypsinated and resuspended in 4-5 ml of DMEM before transferred to a 15 ml tube. Five µl of the resuspended cells were diluted in 495 µl of PBS (1:100) and counted with a Scepter™ 2.0 Cell Counter. Cells were seeded in 6-well plates (1x10^5 cells per well) and grown for 72 hours or in a 12-well plate (2.5x10^5 cells per well) and grown for 24 hours.

The total amount of DNA was 4 µg in 6-wells and 0.5 µg in 12-wells, for each transfection. DNA and Lipofectamine®2000 were both diluted in 250 µl of OPTIMEM. Lipofectamine and OPTIMEM were incubated for 5 minutes at RT. After the incubation, the DNA was transferred to the tube containing Lipofectamine®2000 and incubated for 20 minutes in room temperature. After 20 minutes, the DNA-lipid mix was added to the cells and incubated for 4-6 hours before the cells were washed, and fresh growth media was added. The cells were further grown for 48 hours at 37°C and 5 % CO_2.

4.4.6 Preparation of analytical fractions; medium, cell lysate, and cell pellet

Post-transfection, 1 ml of media was collected and centrifuged at 18 000 x g for 5 minutes at 4°C. The supernatant was collected and analyzed (denoted the ‘medium fraction’). The remaining medium was removed and the cells were washed with PBS and added 150 µl of ice-cold RIPA Lysis Buffer containing protease inhibitors (1 tablet per 10 ml of 1x RIPA buffer). A cell scraper was used to detach the cells from the plate surface. The cells were collected and incubated for 30 minutes on ice, before centrifuged at 18 000 x g at 4°C for 15 minutes. The supernatant was transferred to a new vial and analyzed (denoted the ‘lysate

fraction'). The remaining pellet was washed with 150 µl PBS and centrifuged at 18 000 x g for 15 minutes and 4 °C. This washing step was performed twice. Finally, the pellet was added 100 µl of 4x LDS loading buffer, denatured for 5 minutes at 95°C and analyzed (denoted the 'pellet fraction'). All fractions were stored at -80°C if not analyzed immediately.

4.4.7 Determination of protein concentration

The protein concentration of the lysate fraction was determined by using a Pierce™BCA Protein Assay kit. Four µg of cell lysate were subjected to Western blotting. For the medium and pellet fractions, the volume loaded onto the SDS gel was the same as for the corresponding lysate.

4.5 Western blotting

4.5.1 SDS-PAGE

For SDS-PAGE, the medium and lysate samples were prepared as described in Table 4.8. Before loading, the samples were denatured at 56°C for 15 minutes on a heating block. For the pellet fraction, samples were prepared as described above. Medium (20 µl), lysate (20 µl) and pellet (2-5 µl) samples were loaded on 10 % NuPage Bis-Tris gels (1mm, 10 well) and separated by electrophoresis in a XCell SureLock™ Mini-Cell system. Four µl of Precision Plus Protein Dual Color and 2 µl of Magic Mark XP were used as protein size markers. The gels were run in 1x NuPage MOPS buffer for 15 minutes at 90V and 180 V for 60 minutes.

Table 4.8 Sample preparation for SDS-PAGE

Component	Volume (µl)
Reducing agent 10X	2
LDS 4X	5
Sample/ddH_2O	13
Total volume	**20 µl**

4.5.2 Transfer

After SDS-PAGE, the proteins were transferred from the NuPage Bis-Tris gel to a Polyvinylidene fluorid (PVDF) membrane by Western blotting in an XCell Blot Module chamber according to the manufacturer's instructions. The PVDF membrane was activated in 100 % methanol for 1 minute before use. The blotting was performed at 30 V for 60 minutes

in 1x NuPage transferbuffer with 10 % methanol. Thereafter the PVDF membrane was blocked in 5 % milk (Table 3.10) for 1 hour at room temperature. The membrane was then incubated with primary antibodies diluted in 1% milk o/n at 4°C. The following primary antibodies were used: anti-PAVIRF (1:5000); anti-PRAAHG (1:5000); anti-CEL (1:10 000). As loading control, the primary antibody anti-GAPDH (1:1000) was included. The next day, the membranes were washed 3x5 minutes with PBS-T (0.05%) and incubated with HRP-conjugated secondary antibodies for 1 hour at RT. Secondary antibodies were diluted 1:5000 or 1:10000. Finally, the membranes were washed 1x15 minutes and 3x5 minutes with PBS-T (0.05 %) before developed by using Pierce ECL Plus Western Blotting Substrate kit. The signals were detected with a LAS 1000 imager and quantification of protein bands was performed using the Image Gauge v4.0 software (Fujifilm).

4.5.3 Statistical analysis

The p-value was calculated using a two-tailed T-test. CEL-WT was compared to CEL-TRUNC in the lysate and in the medium fraction. In the pellet fraction, CEL-WT was compared to CEL-INS9, CEL-INS11 and CEL-TRUNC. The significance level was defined as 0.05. To determine the p-value, the statistical calculator from the webpage [http://www.quantitativeskills.com/sisa/statistics/t-test.htm] was used.

4.6 Immunostaining and confocal imaging

Cells were seeded in 12-well plates ($2.5x10^5$ cells per well) on coverslips coated with Poly-L-lysine. After transfection, cells were washed with pre-warmed PBS and fixed in 3% paraformaldehyde (500 µl per 12-well) for 20 minutes.

4.6.1 Staining

After fixation, the cells were washed 3x5 minutes with washing buffer (PBS 0.5 %), before being permeabilized (Table 3.10) for 20 minutes. Next, blocking-buffer (5% goat serum in PBS 0.5 %) was added for 30 minutes before cells were incubated with 40 µl of primary antibody for 2 hours at RT. The primary antibodies used were As20.1 (1:200), anti-PAVIRF (1:5000) or anti-PRAAHG (1:5000) diluted in blocking-buffer. After incubation with primary antibody cells were washed 3x10 minutes in washing buffer. Next, 40 µl conjugated secondary antibodies donkey anti-rabbit Alexa Fluor 488 or anti-mouse Alexa Fluor 488 (1:200 in blocking buffer) was added to each coverslip. Cells were incubated for 1 hour with secondary antibody, before they were washed 3x10 minutes with washing buffer. Finally, the cells were quickly rinsed three times with phosphate buffered saline. Coverslips were mounted on glass slides in 10 µl of Prolong Gold Antifade Solution with 4′,6-Diamidine-2′-phenylindole dihydrochloride (DAPI). Slides were allowed to cure at room temperature overnight and then stored at -20°C.

4.6.2 Confocal imaging

The samples were examined with a Leica TCS SP5 confocal microscope (Leica Microsystems) with a 65×/NA 1.4 Plan-Apochromat oil-immersion objective, ~1.2 Airy unit pinhole aperture and the appropriate filter combinations. Images were acquired with Diode 405 and Argon 488 lasers and processed with Adobe Photoshop CS5 image software (Adobe Systems). Series of images were acquired at 0.25-µm intervals throughout the entire depth of the cells.

4.7 Immunohistochemistry

Formalin Fixed Paraffin Embedded (FFPE) pancreatic tissue sections (3-5µm) were placed onto SuperFrost Plus Adhesion slides and dried overnight at 56°C. The next day, the slides were deparaffinized in xylene for 2x5 minutes and rehydrated with decreasing concentrations of ethanol, 2x5 minutes for each (100 %, 96 % and 80 %) and lastly washed with dH_2O and PBS-T (0.05%). All washing steps were performed on a rocking platform. After washing, the tissue sections were incubated with Tris-EDTA (pH 9) in a pressurized heating chamber at 120 °C for 1 minute and cooled down in running tap water. The tissues were delineated with a Dako pen and incubated with primary antibodies in a humidity chamber overnight at 4 °C. Primary antibodies that were used were anti-PRAAHG (1:5000), anti-PAVIRF (1:5000) and anti-CEL (1:100). The next day, the primary antibody was removed and the tissues were rinsed in dH_2O and washed with PBS-T (0.05%) for 4x15 minutes. The slides were incubated with DAKO peroxidase for five minutes in room temperature, to block endogenous peroxidase activity, rinsed with dH_2O and washed 3x5 minutes in PBS-T (0.05%).

As secondary antibody a kit from Biocare Medical was used. It included MACH3 probe and MACH3 HRP polymer. The probe binds specifically to the primary monoclonal antibody and the polymer horseradish-peroxidase molecule binds specifically to the probe (Biocare Medical). The tissues were incubated with the MACH3 probe for 20 minutes, rinsed in dH_2O and washed in PBS-T (0.05%) for 3x5 minutes. Next, the slides were incubated with MACH3 HRP polymer for 20 minutes before they were rinsed in dH_2O and washed in PBS-T (0.05%) for 3x5 minutes

Visualization was performed by 3,3`-diaminobenzidine (DAB). The sections were stained with hematoxylin for 1 minute in room temperature. Finally, the slides were dehydrated in increasing concentrations of ethanol (80 % – 96 % – 100 %) and xylene, one minute in each. Tissues were mounted in mounting media. To obtain images of the tissue sections, a Leica DM2000LED microscope with the LASV4.8 software or a Leica DMLB microscope with the ZEN 2011 software was used.

5. Results

5.1 Screening for *CEL* insertions in samples from the Pancreas Biobank

The Pancreas Biobank includes both DNA samples and pancreatic tissue sections from most of the patients and thereby makes it possible to study CEL insertions both at DNA and protein levels. To investigate *CEL* insertions at the DNA level we used DNA sequencing.

5.1.1 Amplification of CEL Exon 8-11

Due to the high sequence similarity between *CEL* and the nearby pseudogene *CELP*, one has to be careful when performing PCR analysis on the *CEL* gene. Consequently, before sequencing, we amplified the *CEL* exon 8-11 region using a forward primer (L11F) that binds in intron 7 of *CEL* which is not present in the pseudogene (Fig. 4.1). This PCR product was later used as a template for the sequencing reaction.

To amplify *CEL* exon 8-11, a long-range PCR was performed on genomic DNA samples from the Pancreas Biobank. The PCR products were separated on a 1 % agarose gel to confirm that the amplification had been successful. In Figure 5.1, the molecular bands have the expected length of *CEL* exon 8-11 at approximately 3.8 kb.

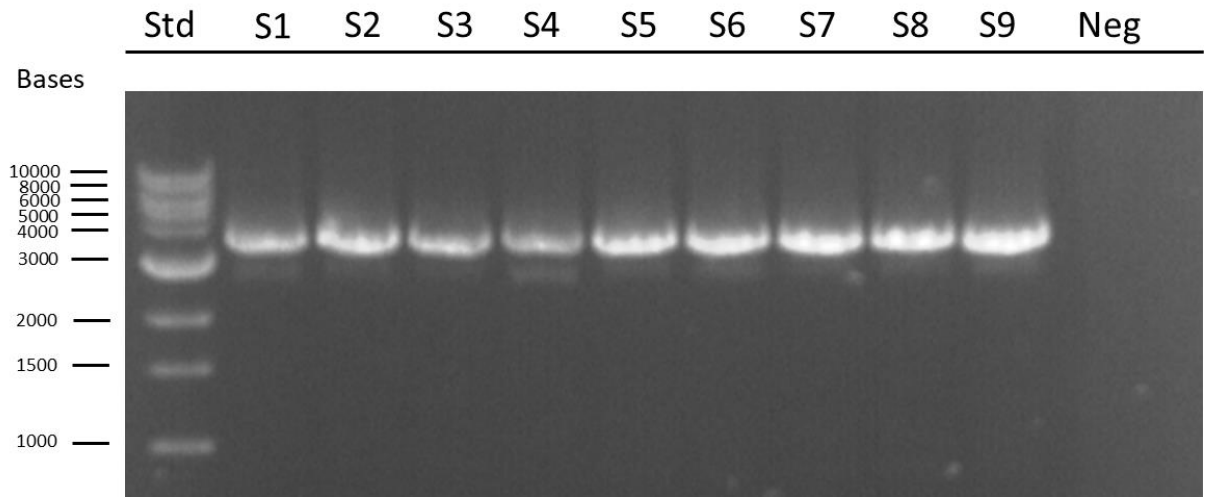


Figure 5.1 *CEL* amplification (exon 8-11) products from long-range PCR visualized by agarose gel electrophoresis. A total of 10 ng genomic DNA from the Pancreas Biobank were loaded. The sizes of the products are about 3.8 kb. The gel picture shows screening of 9 patient samples (lanes S1-S9). Std: Molecular weight standard; Neg: Negative control (with no template).

5.1.2 Sanger sequencing of the CEL VNTR region

To search for insertions in the *CEL*-VNTR region, *CEL* exon 11 was sequenced by Sanger sequencing. The sequence print-outs were evaluated manually. Figure 5.2A shows an excerpt of the VNTR region of CEL where repeat 9-10 is marked. Figure 5.2B displays a close-up image of the same repeat 9-10 and the insertion of a cytosine was found in a region with a high content of cytosines (C). For comparison, repeat 9-10 from a WT sequence is shown in Fig. 5.2C. A total of 50 samples were sequenced and checked for insertions (Table 5.1). Five samples with insertion in repeat 9 (INS9), one sample with an insertion in repeat 10 (INS10) and one sample with an insertion in repeat 12 (INS12) were observed.

The single nucleotide polymorphism (SNP) denoted rs488087 is the only common nucleotide substitution present in the *CEL* VNTR. This variant is a change from a cytosine (C) to a thymine (T) in position 12 in VNTR repeat 2. Of the 50 sequenced samples, 24 subjects with C/C, 21 with C/T and 5 with T/T were observed. This corresponded to a genotype frequency of 0.48 for C/C, 0.42 for C/T, and 0.10 for T/T. The allele frequency for C and T was 0.69 and 0.31, respectively.

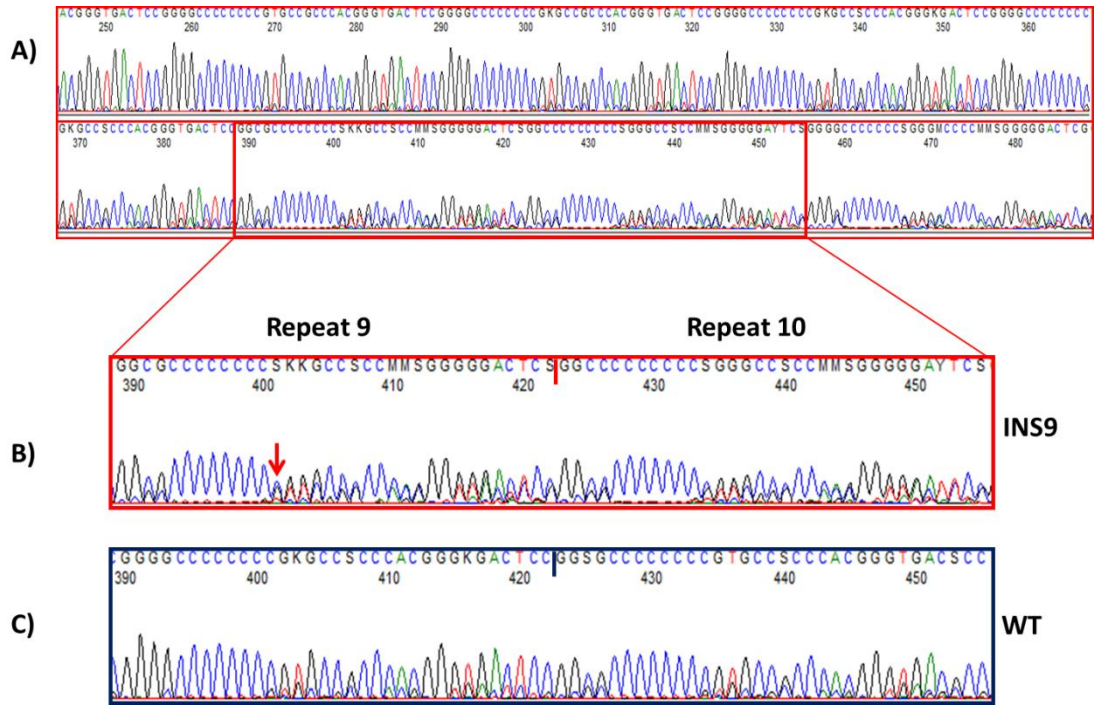


Figure 5.2 Screening for insertions in the VNTR region of *CEL*. Genomic DNA from the Pancreas Biobank was sequenced to look for insertions in the VNTR region of *CEL*. A) An excerpt of a sequence in the VNTR region where repeat 9 and 10 are highlighted. B) Enlarged picture of repeat 9 and 10 in part A. The red arrow points to the site of the insertion of an extra cytosine in repeat 9. C) A WT sequence devoid of cytosine (C) insertion in repeat 9.

Table 5.1 Sequenced genomic DNA samples from the Pancreas Biobank. The table shows a summary of the sequenced genomic DNA samples from the Pancreas Biobank with VNTR length, SNP status in position 12 of VNTR repeat 2, insertion status (marked in red) and the diagnosis of the patients. Samples with VNTR length of 13 are indicated by grey shading. PDAC = pancreatic ductal adenocarcinoma, C= cytosine, T= thymine. INS9 = insertion in repeat 9, INS10 = insertion in repeat 10 and INS12 = insertion in repeat 12.

Sample Id	VNTR length	rs488087 in second VNTR repeat	INS status	Diagnosis
42	14/15	C/C	WT	Benign tumor or other pancreatic disease
43	16/16	C/C	WT	Benign tumor or other pancreatic disease
44	16/16	C/T	WT	PDAC
46	16/16	C/T	WT	PDAC
47	15/16	C/C	WT	PDAC
48	16/16	C/T	INS12	PDAC
49	16/16	C/C	WT	PDAC
50	15/21	C/C	WT	PDAC
51	16/16	C/T	WT	PDAC
52	13/14	C/T	INS9	PDAC
53	16/16	C/T	WT	PDAC
56	15/16	C/C	WT	PDAC
57	13/16	C/C	WT	PDAC
58	14/16	C/C	WT	PDAC
60	16/16	C/T	WT	Benign tumor or other pancreatic disease
61	16/16	C/C	WT	PDAC
62	16/16	C/T	WT	PDAC
63	15/16	C/C	WT	PDAC
64	15/16	C/C	WT	Benign tumor or other pancreatic disease
65	15/16	C/C	WT	PDAC
66	12:15	C/C	WT	Benign tumor or other pancreatic disease
67	16/16	C/C	WT	PDAC
68	16/16	C/C	WT	Benign tumor or other pancreatic disease
69	15/16	C/T	WT	Benign tumor or other pancreatic disease
70	14/15	C/T	WT	PDAC
71	14/16	C/T	WT	Benign tumor or other pancreatic disease
73	16/16	C/C	WT	PDAC
75	16/16	C/T	WT	Benign tumor or other pancreatic disease
76	13/17	T/T	INS9	PDAC
77	16/16	C/T	WT	Benign tumor or other pancreatic disease
78	16/17	C/C	WT	PDAC
79	15/15	C/T	WT	PDAC
80	13/14	C/T	INS9	Benign tumor or other pancreatic disease
82	15/16	C/C	WT	PDAC
83	14/16	C/T	WT	PDAC
84	15/17	C/C	WT	PDAC
85	16/16	C/C	WT	Benign tumor or other pancreatic disease
86	16/16	T/T	WT	PDAC
88	14/15	C/T	INS10	PDAC
89	12:16	C/C	WT	Benign tumor or other pancreatic disease
90	16/16	C/C	WT	PDAC
92	15/15	C/C	WT	Benign tumor or other pancreatic disease
93	12:16	C/C	WT	Benign tumor or other pancreatic disease
94	16/16	C/T	WT	Benign tumor or other pancreatic disease
95	13/16	C/T	INS9	Benign tumor or other pancreatic disease
96	16/16	T/T	WT	PDAC
97	15/16	C/T	WT	PDAC
98	16/16	T/T	WT	Benign tumor or other pancreatic disease
99	16/16	T/T	WT	Benign tumor or other pancreatic disease
100	13/16	C/T	INS9	PDAC

5.1.3 Insertion in repeat 9 is linked to a VNTR length of 13

Among the 50 screened samples, we observed that all samples with an insertion in repeat 9 had a VNTR length of 13 (Table 5.1). To investigate this relationship further, all remaining DNA samples from the Pancreas Biobank with a VNTR length of 13, either homo- or heterozygous, were sequenced (25 samples) (Table 5.2).

Table 5.2 Sequenced genomic DNA samples from the Pancreas Biobank with VNTR length of 13. All genomic DNA samples in the Pancreas Biobank with a VNTR length of 13 were sequenced. The table shows a summary of remaining samples with a VNTR length of 13 plus the six samples from Table 5.1, the SNP status in position 12 in repeat 2, insertion status (marked in red) and the diagnosis of the patient. PDAC = pancreatic ductal adenocarcinoma, C= cytosine, T= thymine. INS9 = Insertion in repeat 9.

Sample Id.	VNTR length	rs488087 in second VNTR repeat	INS status	Diagnosis
3	13/16	T/T	INS9	PDAC
7	13/15	C/T	INS9	Benign tumor or other pancreatic disease
31	13/14	C/C	WT	Benign tumor or other pancreatic disease
52	13/14	C/T	INS9	PDAC
57	13/16	C/C	WT	PDAC
76	13/17	T/T	INS9	PDAC
80	13/14	C/T	INS9	Benign tumor or other pancreatic disease
95	13/16	C/T	INS9	Benign tumor or other pancreatic disease
100	13/16	C/T	INS9	PDAC
101	13/16	C/C	WT	PDAC
103	13/15	T/T	INS9	Benign tumor or other pancreatic disease
106	13/16	C/C	WT	Benign tumor or other pancreatic disease
109	13/15	C/T	INS9	PDAC
113	13/13	T/T	INS9	Benign tumor or other pancreatic disease
128	13/17	T/T	INS9	Benign tumor or other pancreatic disease
132	13/16	C/T	INS9	Benign tumor or other pancreatic disease
140	13/16	T/T	INS9	PDAC
143	13/14/16	C/C	WT	Benign tumor or other pancreatic disease
150	13/16	T/T	INS9	Benign tumor or other pancreatic disease
153	13/13	C/T	INS9	PDAC
164	13/16	T/T	INS9	Benign tumor or other pancreatic disease
174	13/16	C/T	INS9	Benign tumor or other pancreatic disease
188	13/15	C/C	WT	PDAC
209	13/16	C/T	INS9	Benign tumor or other pancreatic disease
210	13/16	C/T	INS9	Benign tumor or other pancreatic disease
220	13/16	C/C	WT	PDAC
236	13/16	C/T	INS9	Benign tumor or other pancreatic disease
240	13/16	C/T	INS9	PDAC
258	13/13	T/T	INS9	Benign tumor or other pancreatic disease
265	13/15	C/T	INS9	PDAC
276	13/16	C/C	WT	PDAC

Of the additional 25 cases with a VNTR length of 13, 18 samples had an insertion in repeat 9. In total, including INS9 cases from Table 5.1 and Table 5.2, 31 samples had a VNTR length of 13. Of these, 23 cases, (74%) had an insertion in the 9th repeat of the VNTR region suggesting a high association between VNTR repeat 13 and the insertion.

Figure 5.3 shows a panel of four sequences (A-D) from repeat 8-9 in the VNTR region of CEL, comparing a WT sequence, a 13 repeat sequence without INS9, and a heterozygous and homozygous sequence with INS9. The homozygous sequence (D) displays a cleaner insertion peak, than the heterozygous sequence (C).

The 13 repeat allele was not linked to rs488087. However, all the INS9 cases had the T allele, which shows a possible association between the T allele and INS9 (Table 5.2).

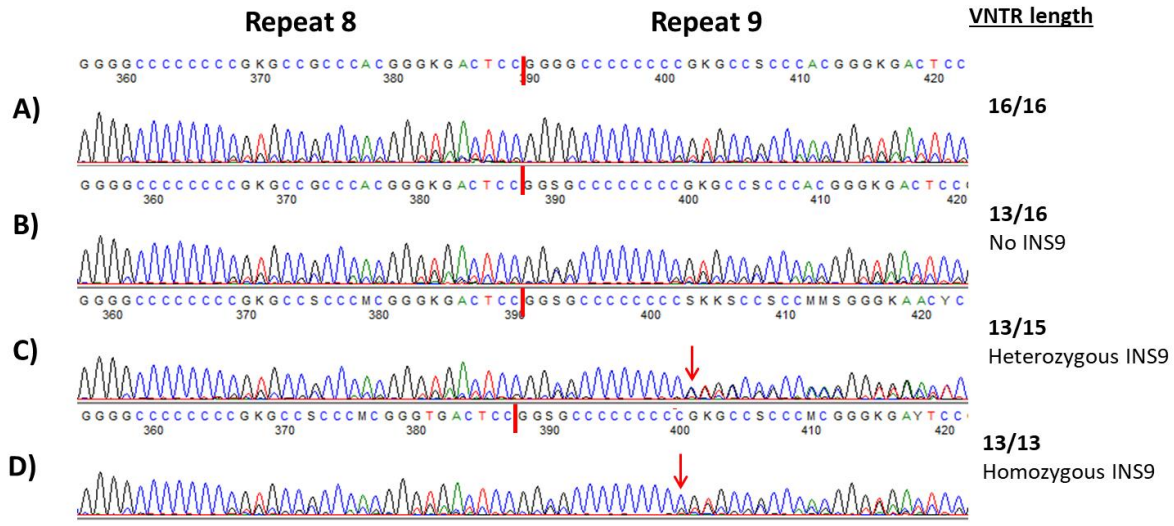


Figure 5.3 Comparison repeat 8 and 9 in sequences with and without INS9. A) WT sequence with a 16/16 VNTR. B) Sequence with 13/16 VNTR, but no insertion. C) Sequence with 13/15 where the red arrow points to the site of insertion. D) Sequence with 13/13 VNTR where the red arrow points to the site of insertion.

5.2 New tools for analyzing insertion variants of CEL

To evaluate and compare the biological function of insertion variants towards the CEL-WT, plasmids expressing different protein variants of CEL (CEL-WT, CEL-INS9, CEL-INS11 and CEL-TRUNC) were used. These plasmids were all a generous gift from (M. E. Lowe and X. Xiao, St. Louis, USA). CEL-TRUNC was included to observe the biological function of CEL without the VNTR region. An empty vector (EV) was included as a negative control. The protein structure of all the CEL variants are presented in Fig. 5.6. All antibodies used in this chapter are specified in Materials (Table 3.11).

5.2.1 Transformation and purification of plasmids expressing different variants of CEL and determination of DNA purity

Plasmids expressing different variants of CEL (CEL-WT, CEL-INS9, CEL-INS11, CEL-TRUNC) or EV were transformed into *E.coli* bacteria. After transformation, the plasmids were purified and analyzed by optical density (OD) measurements to estimate the DNA concentration. In addition, the plasmids were separated on a 1% agarose gel to estimate the quality and confirm that the plasmids were intact (Fig. 5.4).

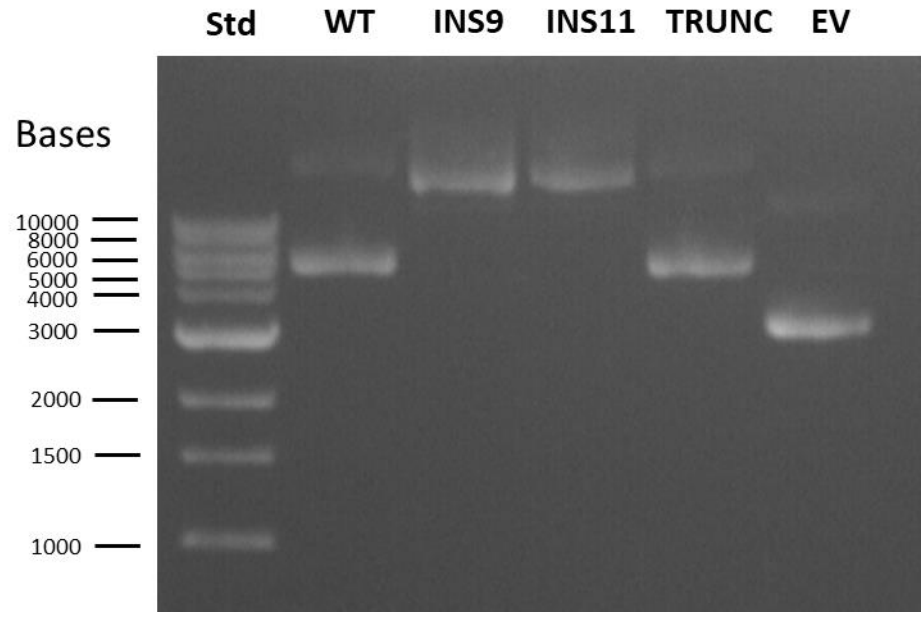


Figure 5.4 Visualization of plasmids expressing different variants of CEL. Plasmids expressing CEL-WT, CEL-INS9, CEL-INS11, CEL-TRUNC and an empty vector (EV) were separated on a 1 % agarose gel stained with ethidium bromide. 600 ng of each plasmid were added to the gel. Std, standard.

5.2.2 Sequencing of plasmids expressing CEL-INS9 and CEL-INS11

To verify that the plasmids expressing CEL-INS9 and CEL-INS11 had the desired mutation in the correct site, the plasmids were sequenced by Sanger sequencing and evaluated by the program SeqScape. Figure 5.5A shows the sequences of repeat 8 and 9 of the CEL-INS9 and CEL-INS11 expressing plasmids. Figure 5.5B shows the sequence of repeat 10 and 11 of the same plasmids. Since the plasmids only differed by the insertion, they served as controls for each other.

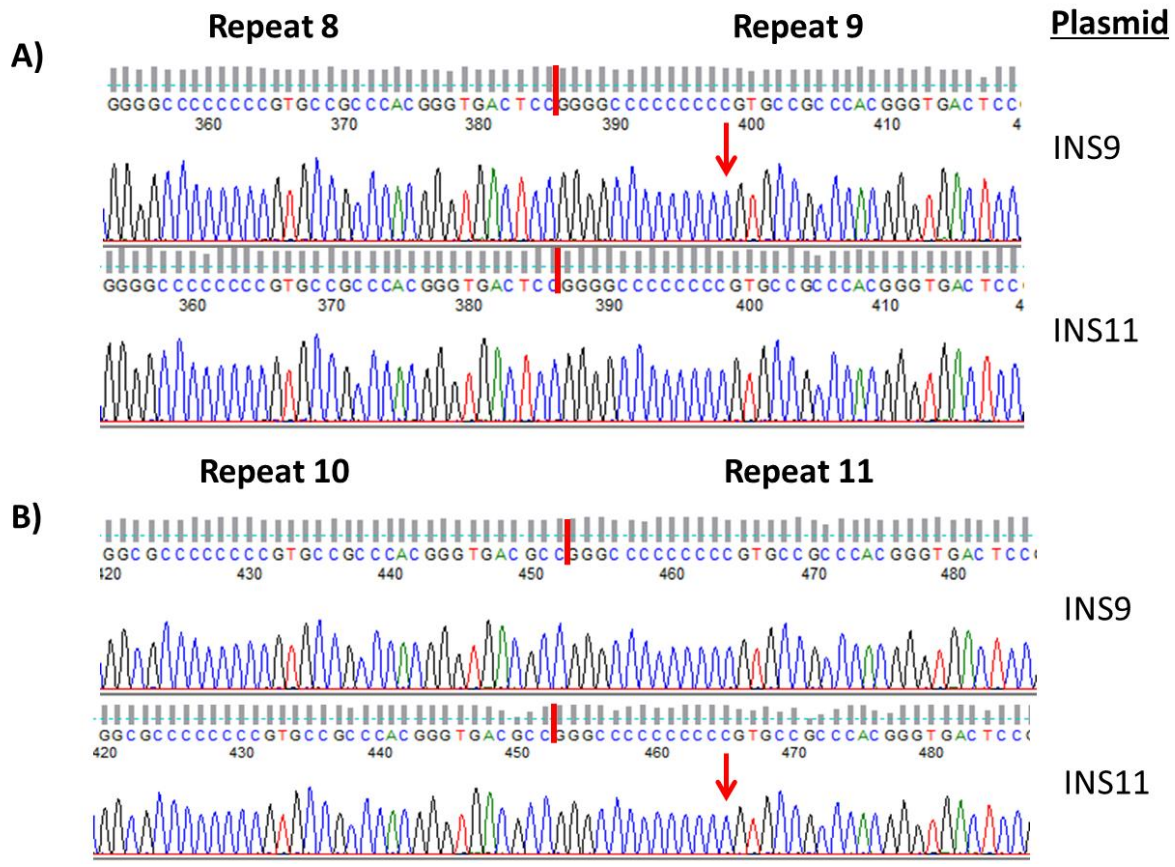


Figure 5.5 Sequencing of plasmids expressing CEL-INS9 and CEL-INS11. Plasmids expressing INS9 and INS11 were sequenced to verify that they had the desired mutation in the right site. A) Repeat 8 and 9 for CEL-INS9 and CEL INS11. The panel shows that INS9 has an insertion of a cytosine in repeat 9 (red arrow), while INS11 has not. B) Repeat 10 and 11 for CEL-INS9 and CEL-INS11. The panel shows that INS11 has an insertion of a cytosine in the 11th repeat (red arrow), while INS9 has not.

5.2.3 Testing of tail-specific antibodies for the CEL-WT and the CEL insertion variants by western blotting

We were provided with antibodies that were specific towards the CEL-WT tail and the CEL-insertion tail (Fig. 5.6). To test these antibodies, HEK293 cells were transiently transfected with plasmids expressing CEL-WT, CEL-INS9, CEL-INS11, CEL-TRUNC or EV (Fig. 5.6). After 72 hours of transfection, the lysate fractions were collected and analyzed by three separate SDS-PAGE and western blots. Immunodetection was performed by a polyclonal CEL-specific antibody (anti-CEL), the CEL insertion-tail specific antibody (anti-PRAAHG) and the CEL WT-tail specific antibody (anti-PAVIRF), Anti-CEL detected all the CEL-variants (WT, INS9, INS11 and TRUNC), anti-PRAAHG detected CEL-INS9 and INS11 and anti-PAVIRF detected the CEL-WT (Fig. 5.7).

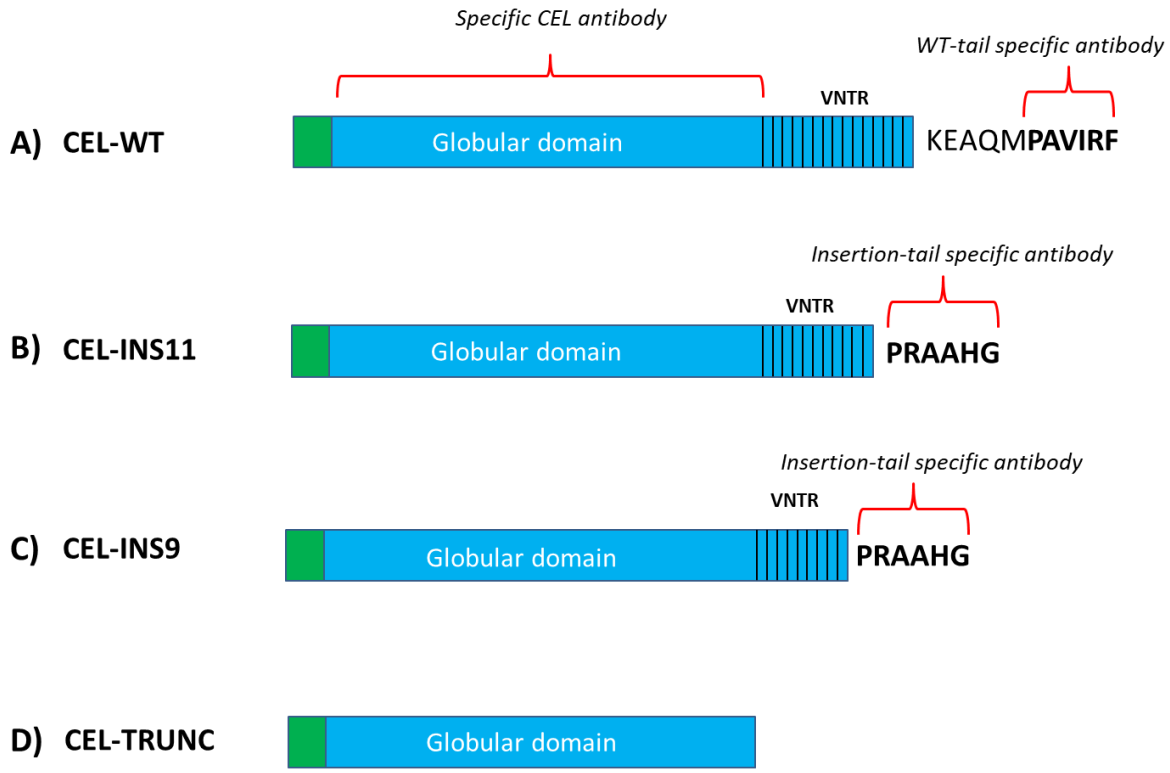


Figure 5.6 Overview of the various CEL varaints tested and their epitopes for the specific antibodies. A) CEL-WT consists of 16 repeats in the VNTR, and the C-terminal ends with the 11 amino acids KEAQMPAVIRF. B and C) CEL-INS11 and CEL-INS9 are insertion variants of CEL and have a truncated C-terminal that ends with the amino acids PRAAHG. D) TRUNC is an artificially created variant of CEL that does not contain the VNTR region. The WT-tail specific antibody binds to the six last amino acids in the C-terminal of CEL-WT, PAVIRF. The insertion-tail specific antibody binds to the six last amino acids in the C-terminal of the insertion variants, PRAAHG. The CEL-specific antibody binds to the globular domain of CEL.

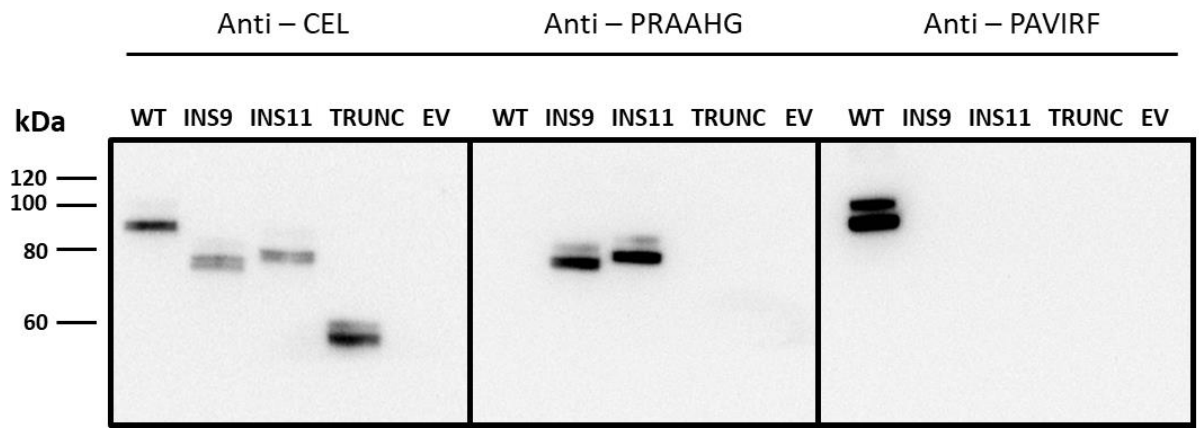


Figure 5.7 Testing of CEL specific antibodies by western blot. HEK293 cells were transiently transfected with different variants of CEL. After transfection, the lysate fractions were collected and further used in SDS-PAGE and western blot with three different antibodies, anti-CEL, anti-PRAAHG and anti-PAVIRF. Anti-PRAAHG is an insertion tail specific antibody, while anti-PAVIRF is a WT-tail specific antibody. For comparison, an antibody towards the globular domain of CEL (anti-CEL) was included.

5.2.4 Testing of CEL-WT and CEL-INS tail-specific antibodies by immunocytochemistry

To observe the reactivity of the specific tail antibodies by immunocytochemistry, HEK293 cells were transiently transfected with plasmids expressing CEL-WT, CEL-INS9, CEL-INS11, CEL-TRUNC or EV. Cells were transfected for 24 hours, before they were fixated and analyzed by immunostaining and confocal imaging. All variants were stained with an anti-CEL-specific antibody, the WT tail-specific antibody (anti-PAVIRF) and the insertion tail-specific antibody, anti-PRAAHG .

The antibody directed towards the CEL globular domain (As20) displayed reactivity to all variants (Fig. 5.8). Further, as expected, the anti-PRAAHG antibody only displayed reactivity to CEL-INS9 and CEL-INS11 (Fig. 5.10). In contrast, the WT-tail specific antibody that should only have reactivity to the HEK293 cells expressing CEL-WT also showed reactivity to both INS-variants and CEL-TRUNC, which lacks the WT-tail (Fig. 5.9A). However, the positive staining in Figure 5.9A may be misleading, since the positive reactions only occurred in very few cells. Thus, Figure 5.9B is included to give a more representative image.

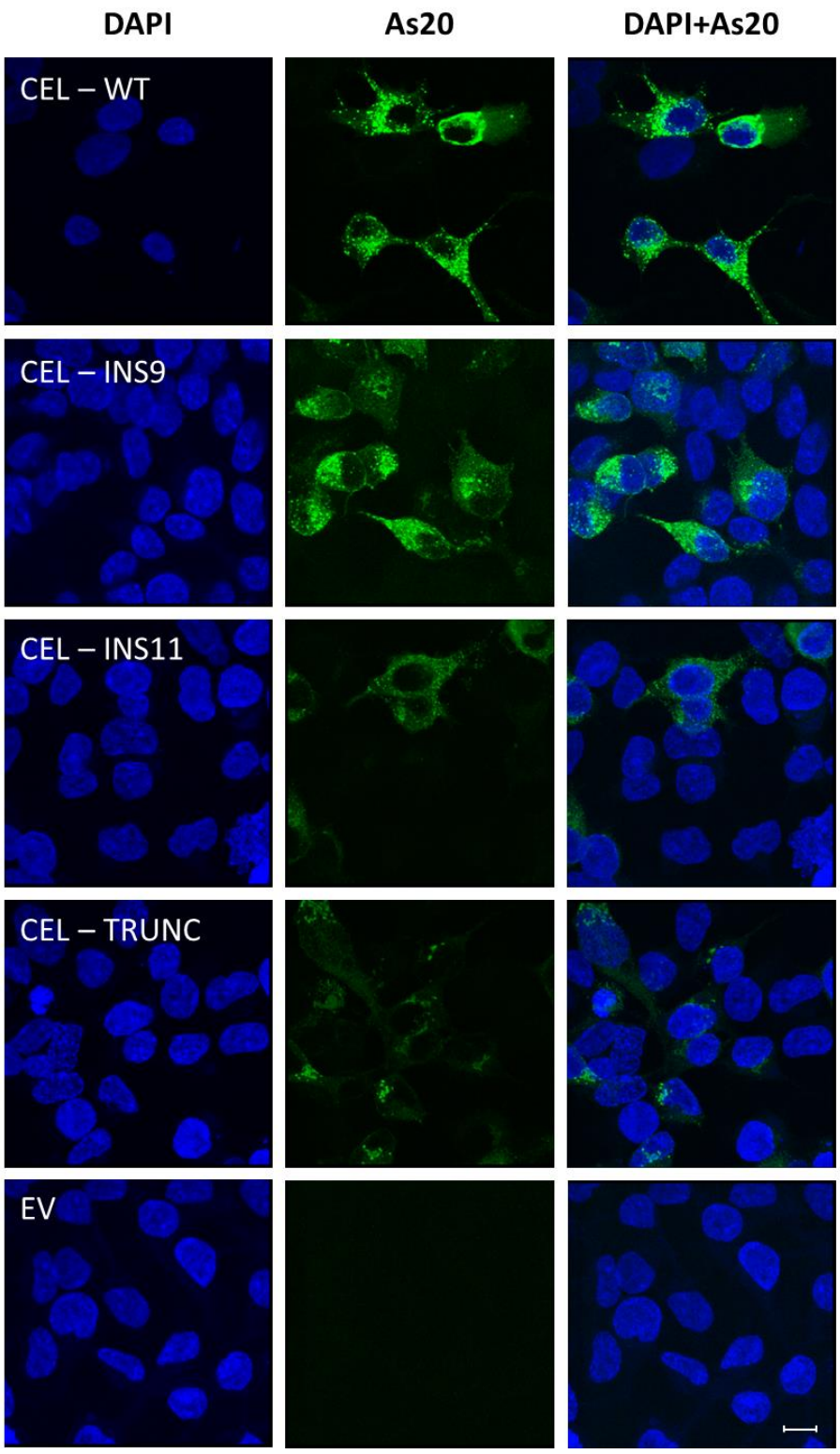
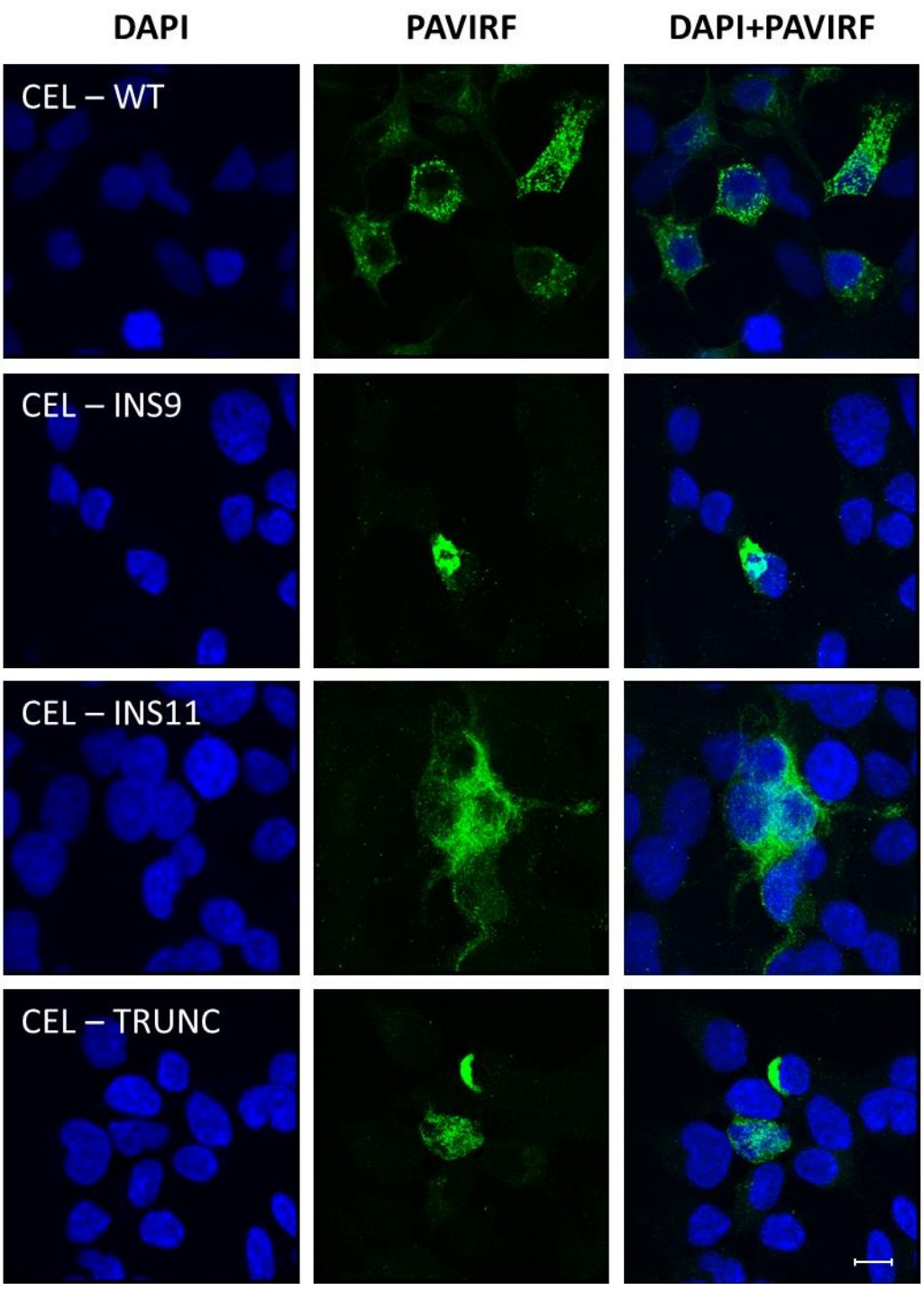


Figure 5.8 CEL immunostaining using an antibody directed towards the CEL globular domain (anti-As20). HEK293 cells were transiently transfected with CEL-WT, CEL-INS9, CEL-INS11 and CEL–TRUNC. An empty vector (EV) was included as control. Post-transfection, the cells were subjected to immunostaining using the CEL specific As20-antibody and secondary antibody AlexaFluor-488 (green). The chromosome counterstain (DAPI) is shown in blue. Each image represents a maximum intensity projection of a z-stack taken through the entire depth of the cell. The scale bar represents 10 μm.

A)



B)

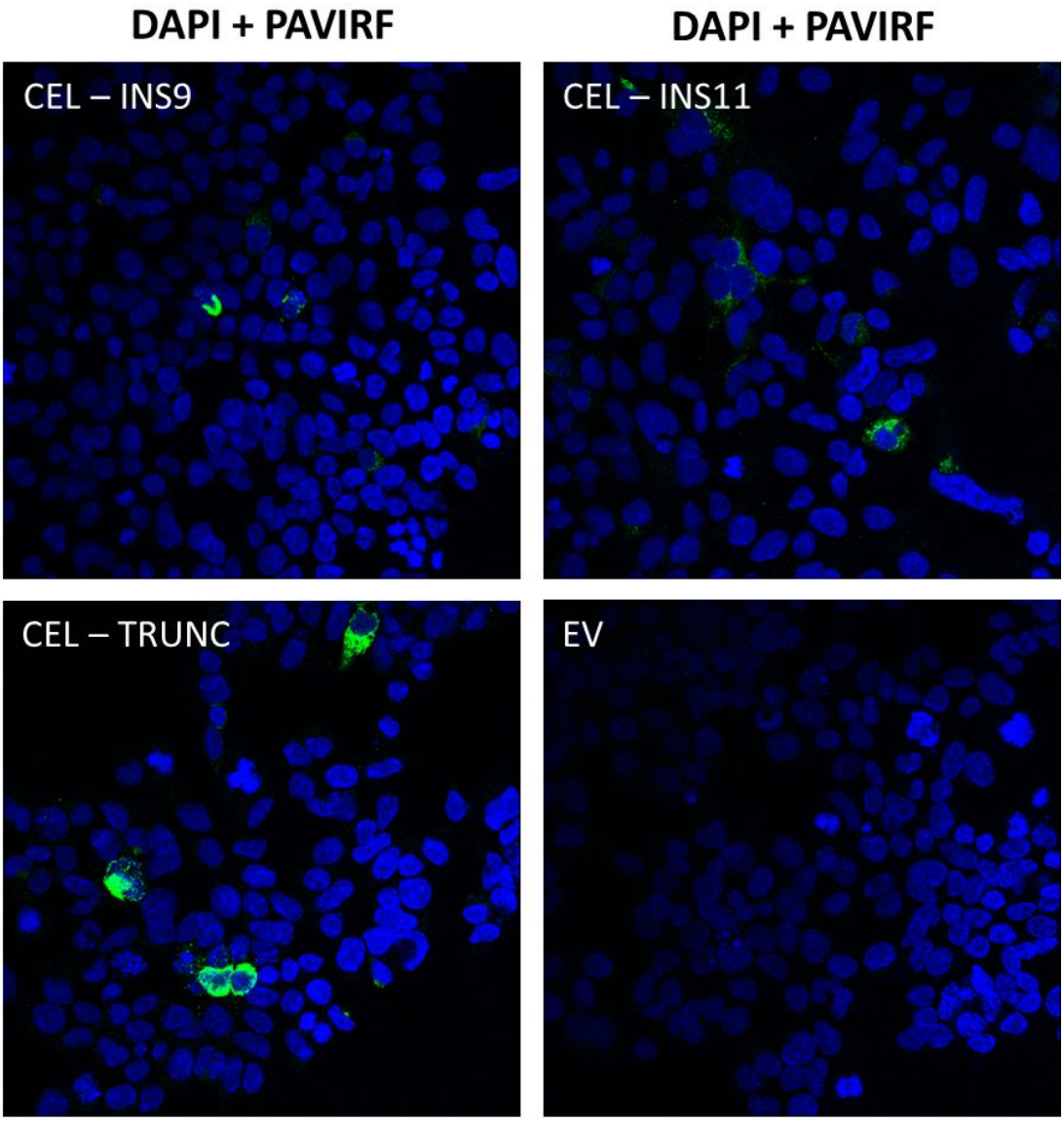


Figure 5.9 CEL immunostaining using a WT-tail specific antibody (anti-PAVIRF). A) HEK293 cells were transiently transfected with CEL-WT, CEL-INS9, CEL-INS11 and CEL – TRUNC. Empty vector (EV) was included as a negative control. 48 hours post-transfection, the cells were subjected to immunostaining using the Anti-PAVIRF antibody. As a secondary antibody, AlexaFluor-488 (green) was used. The chromosome counterstain (DAPI) is shown in blue The scale bar represents 10 μm. B) Overview images of CEL-INS9, CEL-INS11, CEL-TRUNC and EV stainings. Each image represents a maximum intensity projection of a z-stack taken through the entire depth of the cell.

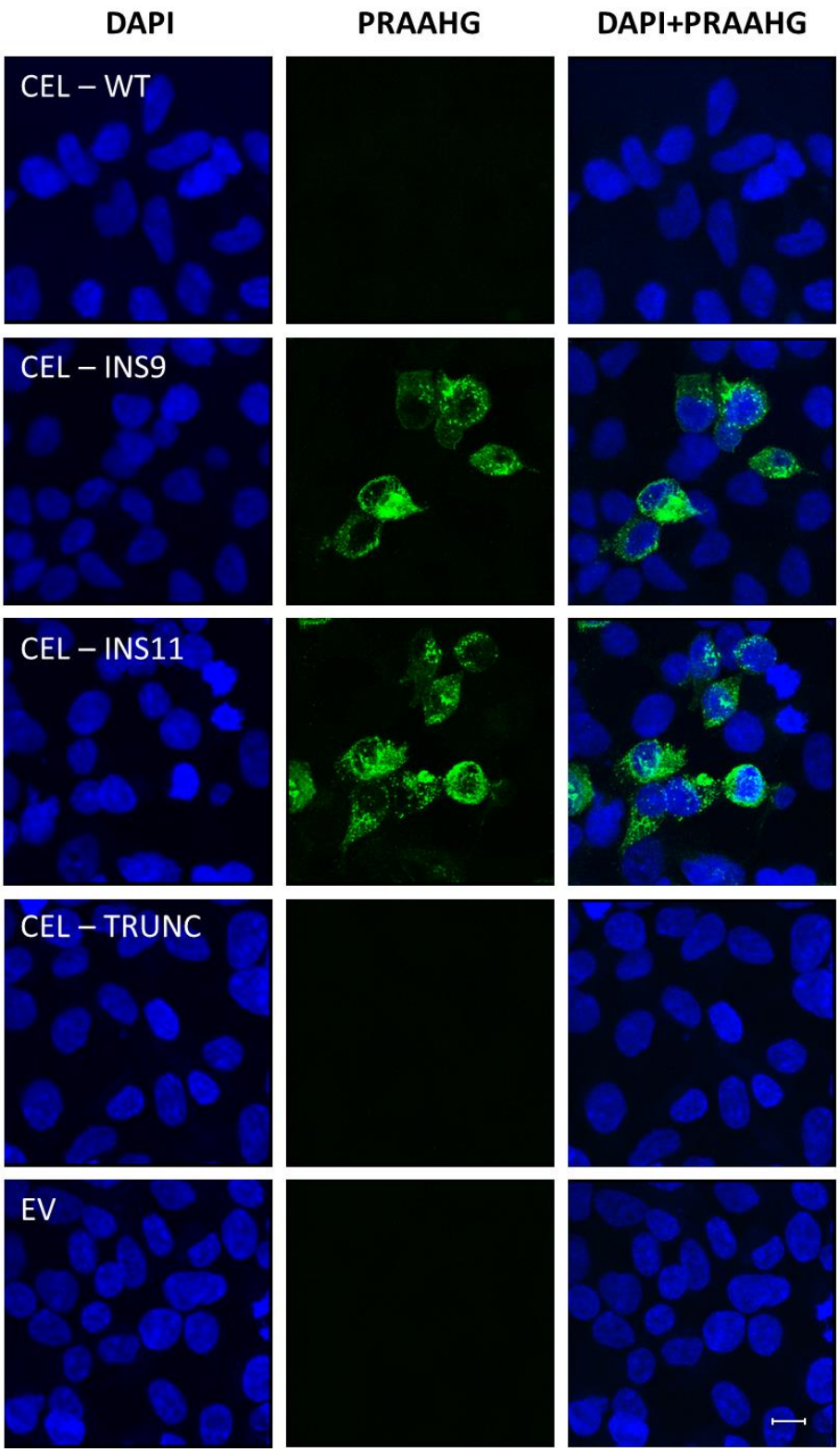


Figure 5.10 CEL immunostaining using an INS-specific tail-antibody (anti-PRAAHG). HEK293 cells were transiently transfected with CEL-WT, CEL-INS9, CEL-INS11 and CEL-TRUNC. Empty vector (EV) was included as a negative control. 48 hours post-transfection, the cells were subjected to immunostaining using the anti-PRAAHG antibody. As a secondary antibody, AlexaFluor-488 (green) was used. The chromosome counterstain (DAPI) is shown in blue. Each image represents a maximum intensity projection of a z-stack taken through the entire depth of the cell. The scale bar represents 10 μm.

5.3 Testing tail-specific antibodies on human pancreatic tissues

We next wanted to determine whether the tail-specific antibodies also could be used to detect CEL-variants in human pancreatic tissue. Formalin fixed paraffin embedded sections from three different patients from the Pancreas Biobank were selected for testing.

5.3.1 Testing CEL tail-specific antibodies with immunohistochemistry on pancreatic ductal adenocarcinoma cases

Immunohistochemistry were performed on formalin fixed paraffin embedded morphological normal appearing tissues from patients with pancreatic ductal adenocarcinoma. This was performed to observe the reactivity of the CEL-insertion (anti-PRAAHG) and the CEL-WT (anti-PAVIRF) tail-specific antibodies. Figure 5.11 show three samples from Table 5.2 (153, 240 and 220), all stained with anti-PRAAHG, anti-PAVIRF and anti-CEL. All samples showed reactivity to anti-CEL and anti-PAVIRF. Number 153 and 240 were insertion-positive and displayed reactivity to anti-PRAAHG. Number 220 was insertion-negative and was expected to be negative, but interestingly still had some isolated areas that were anti-PRAAHG positive.

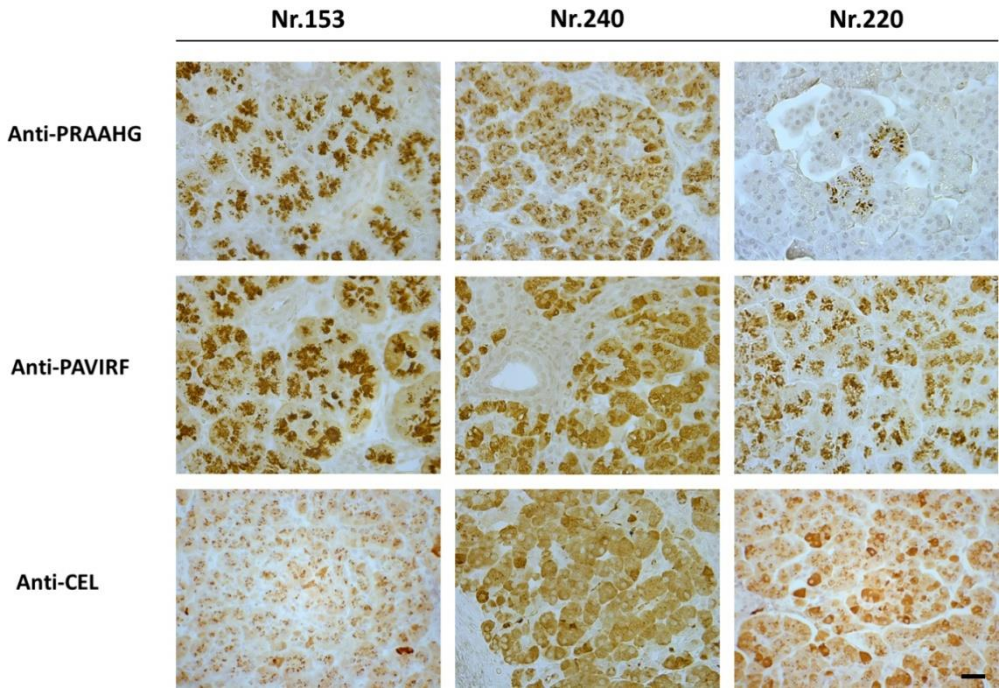


Figure 5.11 Reactivity of pancreas tissues to CEL specific antibodies. Morphological normal appearing tissue sections from pancreas cancer patients were stained with anti-CEL (globular domain-specific, Sigma), anti-PAVIRF (WT tail-specific) and anti-PRAAHG (insertion tail-specific). Positive CEL staining is shown in brown. Nr. 153, 220 and 240 represents three different PDAC cases. Sample 153 and 220 is insertion positive and 240 is insertion negative. Scale bar is 20 µm.

5.3.2 Testing of CEL tail-specific antibodies with immunohistochemistry on pancreatic tissue sections from patients with non-cancerous disease

We observed that sample nr. 220 (Fig. 5.11) had restricted positive areas with anti-PRAAHG. However, all the cases in Figure 5.11 were PDAC cases and we wondered if this observation was specific for cancer patients. To observe the reactivity of non-pancreatic cancer cases with the tail-specific antibody anti-PRAAHG, immunohistochemistry was performed on four formalin fixed paraffin embedded human tissues from pancreas. The four non-pancreatic cancer cases all had clusters of cells or single cells that showed reactivity to anti-PRAAHG (Fig. 5.12). An insertion-positive pancreatic cancer case was included as a positive control (Nr. 153, Table 5.2).

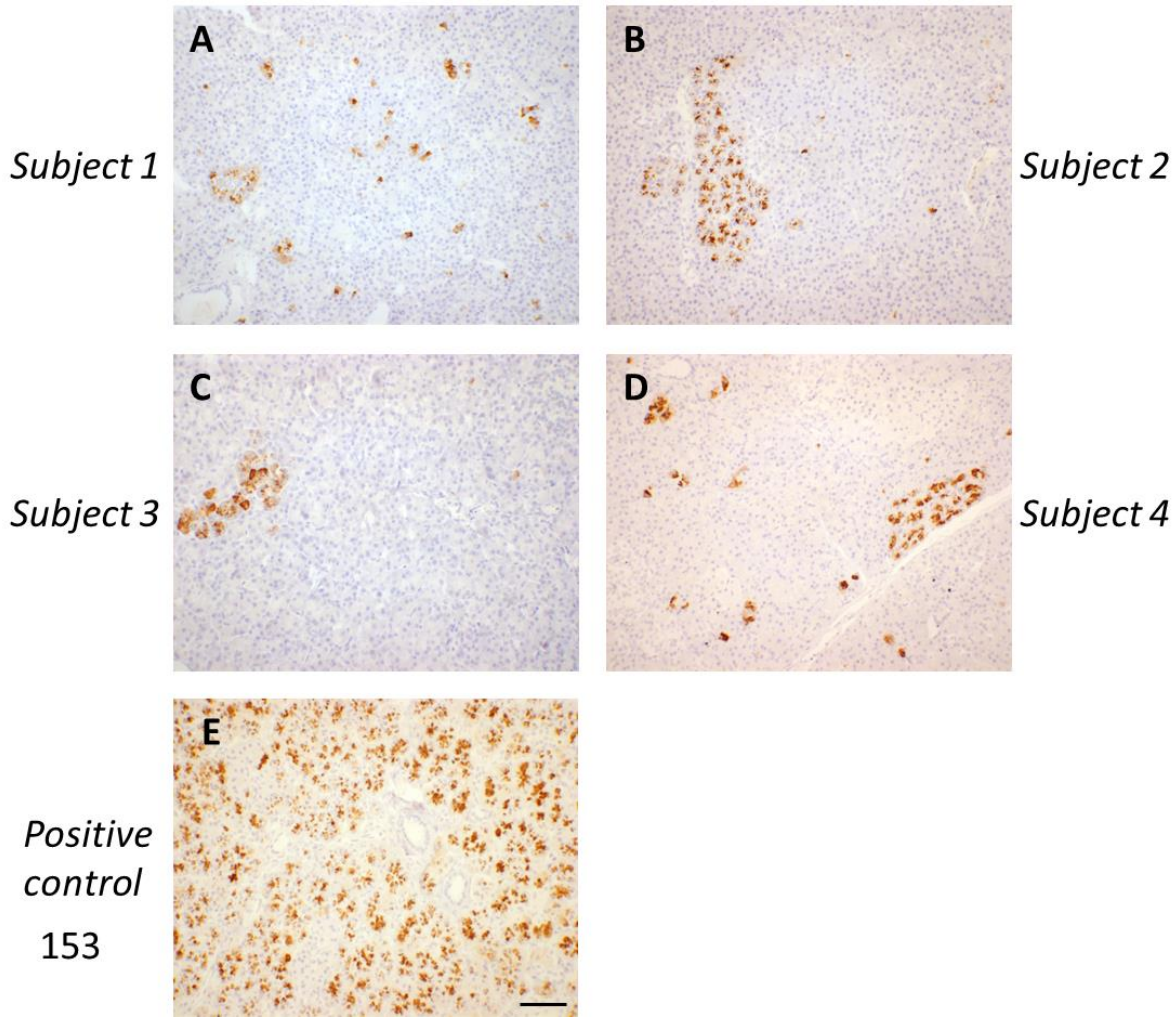


Figure 5.12 Reactivity of pancreas tissues in non-cancer patients with the insertion tail-specific antibody. Non-pancreatic cancer tissues (subject A-D) were stained with the insertion tail-specific antibody (anti PRAAHG) (brown). An insertion-positive pancreatic cancer case (number 153 in table 5.2) was included as a positive control. Scale bar is 50 µm.

5.4 The effect of insertions in the CEL-VNTR region on protein secretion and aggregation

Finally, we wanted to investigate the impact of the insertion variants in a cellular model system to observe if the insertion variants had a potential pathogenic effect compared to the CEL-WT.

HEK293 cells were transiently transfected with plasmids expressing different variants of CEL, namely CEL-WT, CEL-INS9, CEL-INS11 and CEL-TRUNC (Fig. 5.6). CEL-TRUNC was included to observe the effect of protein secretion and aggregation with an absent VNTR region. Cells transfected with an empty vector (EV) was included as a negative control. After 48 hours of transfection, cell medium, lysate and pellet fractions were isolated and analyzed by SDS-PAGE and western blotting. GAPDH was used as a loading control. Immuno-detection was performed using a polyclonal CEL-specific antibody (Xiao et al., 2016).

Three independent experiments were performed (Fig. 5.13A) and quantification analysis of the Western blots results are shown in Figure 5.13B. The diagram shows that INS9 and INS11 are approximately equally detected in the lysate and the medium, but in the pellet there is a higher content of INS9 than of INS11 and CEL-WT. CEL-TRUNC was detected in the lysate, almost nothing in the medium and the highest content of CEL-TRUNC was observed in the pellet. The differences observed between the CEL variants were not statistical significant.

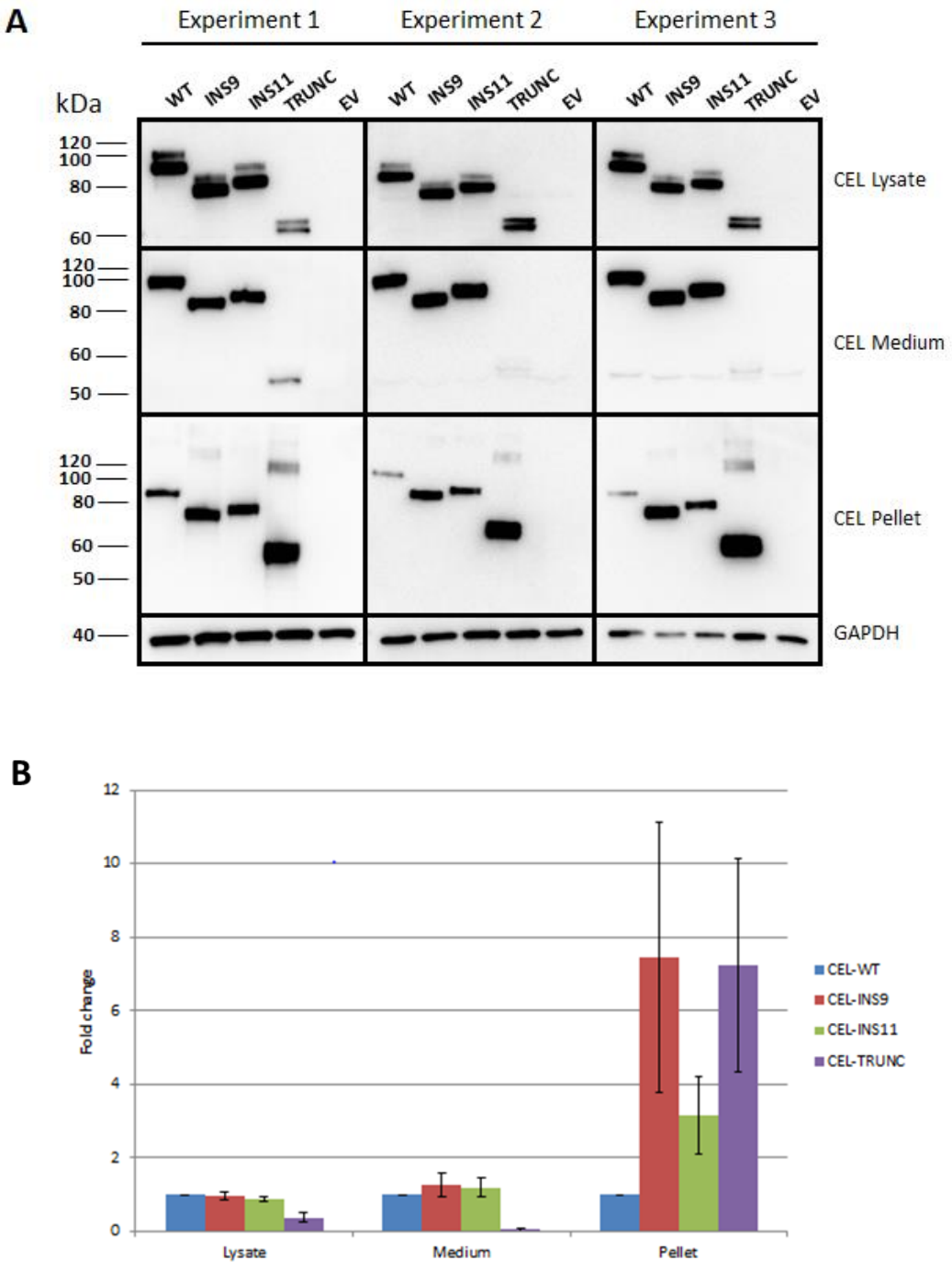


Figure 5.13 The impact of CEL insertion variants on protein secretion and aggregation compared to CEL-WT and CEL-TRUNC. HEK293 cells were transiently transfected with different variants of CEL: CEL-WT, CEL-INS9, CEL-INS11 and CEL-TRUNC. Post-transfection, medium, lysate and pellet fractions were collected and analyzed by SDS-PAGE and western blot. A) Three western blots, representing three individual experiments. GAPDH was used as loading control. The chemiluminescence exposure was 10 seconds for all blots. B) Quantification of the intensity of each band in the three individual western blot experiments. The intensity of the CEL bands was adjusted to GAPDH and normalized towards the CEL-WT. The data represents the mean of the normalized values and the error bars are standard error of the mean (SEM) of three individual experiments.

6. Discussion

Diseases that are associated with carboxyl ester lipase (CEL) are mainly related to alterations occurring in the variable number of tandem repeat (VNTR) region. Disease progression in HIV-1 has been linked to variations in VNTR length (Stax et al., 2012), a longer VNTR has been associated with alcoholic liver cirrhosis (Fjeld et al., 2016) a CEL-deletion variant (CEL-HYB) gives an increased risk for developing chronic pancreatitis (Fjeld et al., 2015). The strongest association is seen for MODY8, a syndrome of diabetes and exocrine dysfunction. The syndrome is caused by a single base pair deletion in the *CEL* VNTR region and results in a somewhat shorter protein with a completely changed C-terminal due to an altered reading frame and a premature stop codon (Ræder et al., 2006).

One base-pair insertions in the *CEL* VNTR are quite common, the combined allele frequency in normal controls of Northern European descent is around 0.07 (Ræder et al., 2006). Such insertions also result in a premature stop codon and a truncated protein. However, compared to the MODY8-causing deletion, the stop-codon arises in the same repeat where the insertion occurred. Thus, an insertion will lead to a short downstream segment and a truncated protein whereas a deletion causes a long and new protein sequence which easily aggregates (Ræder et al., 2006, Johansson et al., 2011, Xiao et al., 2016). Insertions in the VNTR have been suggested to associate with an increased risk of exocrine dysfunction in diabetes patients (Ræder et al., 2006) and with pancreatic cancer (Martinez et al., 2016), but their functional properties have not been studied.

In this project, we aimed to know more about the CEL VNTR insertions variants both at the DNA and protein level. We genotyped the VNTR region in pancreatic cancer patients and analyzed the localization and frequency of *CEL* insertions. Tail-specific antibodies for CEL-WT and CEL insertion variants were tested for western blotting, immunocytochemistry and immunohistochemistry. Moreover, the effect of CEL VNTR insertions on protein secretion and aggregation was studied.

6.1 Insertions in CEL VNTR repeat 9 are linked to a VNTR length of 13

Initially, we searched for insertion variants in the *CEL* gene in a small material of pancreatic cancer patients. We wanted to determine the location of the insertions within the VNTR and their frequency. We screened 50 samples and found five cases with insertion in repeat 9, one case with insertion in repeat 10 and one case with insertion in repeat 12. Furthermore, from our initial screening (Table 5.1), we observed that all five cases with insertions in repeat 9 had a VNTR length of 13 repeats on one allele. This could indicate that insertion in repeat 9 is linked to the 13-repeat allele.

To investigate this possible association further, we sequenced the CEL VNTR of all remaining samples in the Pancreas Biobank with a VNTR length of 13 repeats (N=25). Of these new cases, 18 had an insertion in repeat 9. In total, including INS9 cases from Table 5.1, 23 of 31 cases with a 13 repeat VNTR length had insertion in repeat 9 (Table 5.2). For samples that were heterozygous for a VNTR length of 13 (e.g. 13/16), we did not know whether the insertion was located on the allele with 13 repeats or on the other allele. However, three of the 31 samples with a VNTR length of 13 were homozygous (13/13) and all of them were also homozygous for the insertion in repeat 9.

Taken together**,** our findings strongly suggest the one-base pair insertion in the *CEL* VNTR repeat 9 is located on an allele with 13 repeats. Nevertheless, we observed that not all cases with the 13 repeat allele had INS9. To decide if the observed correlation between VNTR-length and INS9 has a biological impact, further investigations are needed.

6.2 Is the SNP rs488087 linked to pancreatic ductal adenocarcinoma?

In a previous study, the variant rs488087 was suggested to result in an increased risk for developing pancreatic ductal adenocarcinoma (PDAC) (Martinez et al., 2015). However, that study used a patient cohort of only 30 patients, which is far too small to make such a suggestion. The same argument can be used for the current study: 50 cases is not a large enough patient cohort to conclude that any SNP is associated with an increased risk for developing PDAC. In addition, a control cohort is needed to see how common this SNP is in the normal population compared to patients with PDAC. However, according to the SNP database from NCBI, [https://www.ncbi.nlm.nih.gov/snp/] the allele frequency in Europe for SNP rs488087 is 0.69 for the C-allele and 0.31 for the T-allele. These are exactly the same allele frequencies found in our study. Thus, from the current data, we obtained no indication that rs488087 is associated with pancreatic cancer risk in the European population.

6.3 The specificity of the tail-specific antibodies

A CEL-WT sequence has a C-terminal that ends with the amino acid sequence PAVIRF. If an insertion occurs in the VNTR, it will result in a premature stop codon, a truncated protein and a C-terminal sequence that ends with PRAAHG. In a recent study, the authors produced antibodies that were specifically directed to these two C-terminal ends, PAVIRF and PRAAHG and used them for Western blotting, immunocytochemistry and immunohistochemistry (Martinez et al., 2016).

When we used the same antibodies for western blotting (Fig. 5.7), the tail-specific antibodies were highly specific and corresponded to the results published by Martinez and co-workers (Martinez et al., 2016). Surprisingly, however, when performing immunofluorescence staining on transiently transfected HEK293 cells, the CEL-WT tail-specific antibody (anti-PAVIRF) exhibited reactivity also towards the insertion variants and CEL-TRUNC (Fig. 5.9a). In the (Martinez et al., 2016) study, unspecific binding was not observed for the anti-PAVIRF antibody by immunocytochemistry. One possible reason could be that we use different protocols. Further, it should be noted that by immunocytochemistry the proteins are detected in a more native form compared to western blot analysis where the proteins are denatured. Thus, antibodies could work differently in immunocytochemistry and western blotting.

Interestingly, when performing immunohistochemistry on PDAC sections (Fig. 5.11) we observed that one individual without a *CEL* VNTR insertion variant in the germline had small areas with reactivity towards the anti-PRAAHG antibody. Martinez et al. reported the same for eight patients with PDAC, further, they hypothesized that insertions in the VNTR region of CEL could be associated with pancreatic cancer and appear as somatic mutations in early development of PDAC (Martinez et al., 2016). Since analysis of the anti-PRAAHG antibody on pancreatic tissues without cancer was left out of their study we decided to test it on pancreatic tissues from four relatively young patients without PDAC (but with other pancreatic disease). We observed small anti-PRAAHG positive areas or single cells also in normal pancreatic tissue - as in the PDAC case without insertion in the germline (Fig. 5.11 and 5.12). This may indicate that insertions occur also as somatic mutations in individuals without PDAC. However, unspecific anti-PRAAHG antibody binding cannot be ruled out.

6.4 Is the insertion mutations pathogenic?

Two clearly pathogenic variants of CEL have been reported: CEL-MODY and CEL-HYB (Ræder et al., 2006, Fjeld et al., 2015). CEL-MODY is caused by a single base pair deletion which causes an altered reading frame that results in a protein that very easily forms intracellular and extracellular aggregates (Johansson et al., 2011). CEL-HYB has three VNTR repeats originating from its neighboring pseudogene. When CEL-HYB was expressed in a cellular model system, proteins tend to accumulate inside the cell to a much higher degree than the normal protein. In addition, CEL-HYB showed impaired secretion (Fjeld et al., 2015).

In this study, we wanted to investigate the possible pathogenic effects of CEL insertion variants. By cellular fractionation of HEK293 cells we observed that the insertion variants were similarly detected in the cell lysate and the medium as the CEL-WT protein (Fig. 13b). Interestingly, we observed somewhat more CEL-INS9 than CEL-INS11 in the pellet fraction. Since CEL-INS9 has a shorter C-terminal than INS11, the ability to form insoluble protein aggregates may increase with shorter VNTR-lengths. To support this, almost nothing of the CEL-TRUNC protein was secreted, and instead this variant was highly abundant in the pellet fraction. Also, from previous studies on the known pathogenic CEL variants, both amino acid composition and length of the C-terminal tail have shown to be crucial for the cellular properties of CEL (Torsvik et al., 2014, Fjeld et al., 2015).

Here, we used HEK293 cells. An advantage is that they are secretory cells from a human origin and do not express CEL endogenously (which otherwise may have disturbed the interpretation of our results). However, the HEK293 cells are human embryonic kidney cells, and may not give a correct image of how insertion variants of CEL behave in their natural environment. Thus, it would be more representative to use acinar cells. Two acinar cell lines are commercially available: rat acinar cell (ARJ42J) and mouse acinar cells (266-6). However, these cell lines are difficult to culture and are biologically different from human acinar cells. Unfortunately, no human cell line is available to date.

Based on our cellular studies, insertions may have an effect of the cellular properties of the protein. In addition, we observed that the carrier frequency of *CEL* insertions in a Norwegian pancreas biobank (N=50) was 14 %. However, due to the high carrier frequency of insertions in the general population (about 16 %; (Ræder et al., 2006) and the fact that we detected possible somatic CEL insertions in patients without PDAC, we cannot confirm that CEL insertions are associated with pancreatic cancer. Thus, it is yet not possible to use CEL insertions as an early detection tool for PDAC, as suggested by Martinez et al (Martinez et al., 2016) However, our study is based on *CEL* insertions in VNTR repeat 9, 10 and 12. Investigations of insertion in other VNTR alleles are also needed to completely rule of any possible link between CEL VNTR insertions and pancreatic cancer.

7. Conclusions

The overall aim for this study was to understand the biological impact of insertion variants in the *CEL* gene. Based on our findings the following conclusions can be drawn:

- We have established a strategy for how to analyze CEL insertion variants both at the DNA and protein level
- In samples from a Norwegian pancreas biobank, insertions in *CEL* VNTR repeat 9 are associated with a VNTR length of 13
- There is no indication that the *CEL* SNP rs488087 is associated with pancreatic cancer risk in the European population
- CEL insertion variants cannot yet be used as biomarkers to detect somatic mutations in pancreatic cancer
- Compared to normal CEL, proteins with an insertion in VNTR repeat 9 or 11 may have a higher tendency to accumulate in the insoluble pellet fraction when expressed in HEK293 cells

8. Future perspectives

The VNTR region is a critical region which should be further investigated to gain a better understanding of how CEL is involved in pancreatic disease. To analyze CEL VNTR insertions in more depth, one should focus on the following studies:

- Analyze larger patient and control cohort - to further study insertions in the VNTR region and to investigate if insertions could be located in other repeats than 9, 10 and 12
- Sequence DNA from the patients without PDAC that displayed reactivity to anti-PRAAHG to observe if the insertions were germline or somatic
- Test the tail-specific antibodies in protein extracts from pancreatic juice isolated from patients with and without insertions in the germline
- Study the effect of a CEL variant with an insertion in a repeat number lower than 9, to investigate if this would have stronger impact on protein properties than what observed for INS9
- Transfect CEL insertion variants into acinar cells to analyze them in a more biologically representative environment.

9. References

AABAKKEN, L. 2016. *Bukspyttkjertelbetennelse* [Online]. Store medisinske leksikon Available: https://sml.snl.no/bukspyttkjertelbetennelse [Accessed 10.02.18 2018].

ABOUAKIL, N. & LOMBARDO, D. 1989. Inhibition of human milk bile-salt-dependent lipase by boronic acids. Implication to the bile salts activator effect. *Biochimica et Biophysica Acta* 1004**,** 215-220.

ABOUAKIL, N., MAS, E., BRUNEAU, N., BENAJIBA, A. & LOMBARDO, D. 1993. 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. *The Journal of biological chemistry,* 268**,** 25755-25763.

AMERICAN DIABETES ASSOCIATION 2017. Classification and diagnosis of diabetes. *Diabetes Care,* 40**,** S11.

BENGTSSON-ELLMARK, S. H., NILSSON, J., ORHO-MELANDER, M., DAHLENBORG, K., GROOP, L. & BJURSELL, G. 2004. Association between a polymorphism in the carboxyl ester lipase gene and serum cholesterol profile. *European Journal of Human Genetics,* 12**,** 627-632.

BERTELSEN, B. I. 2014. *Menneskets sykdommer*, Gyldendal Akademisk.

BIOCARE MEDICAL. *MACH 3 Rabbit HRP Polymer Detection* [Online]. Available: https://biocare.net/product/mach-3-rabbit-hrp-polymer-detection/ [Accessed 05.05 2018].

BLÄCKBERG, L. & HERNELL, O. 1983. Further characterization of the bile salt-stimulated lipase in human milk. *FEBS Letters,* 157**,** 337-341.

BLÄCKBERG, L., LOMBARDO, D., HERNELL, O., GUY, O. & OLIVECRONA, T. 1981. Bile salt-stimulated lipase in human milk and carboxyl ester hydrolase in pancreatic juice. *FEBS Letters,* 136**,** 284-288.

BRUNEAU, N., LOMBARDO, D. & BENDAYAN, M. 1998. Participation of GRP94-related protein in secretion of pancreatic bile salt-dependent lipase and in its internalization by the intestinal epithelium. *J. Cell Sci.,* 111**,** 2665-2679.

BRUNEAU, N., NGANGA, A., FISHER, E. A. & LOMBARDO, D. 1997. O-Glycosylation of C-terminal tandem-repeated sequences regulates the secretion of rat pancreatic bile salt-dependent lipase. *The Journal of biological chemistry,* 272**,** 27353-27361.

BURGOYNE, R. D. & MORGAN, A. 2003. Secretory granule exocytosis. *Physiol. Rev.*

COMTE, B., FRANCESCHI, C., SADOULET, M. O., SILVY, F., LAFITTE, D., BENKOEL, L., NGANGA, A., DANIEL, L., BERNARD, J. P., LOMBARDO, D. & MAS, E. 2006. Detection of bile salt-dependent lipase, a 110 kDa pancreatic protein, in urines of healthy subjects. *Kidney International,* 69**,** 1048-1055.

DALVA, M., EL JELLAS, K., STEINE, S. J., JOHANSSON, B. B., RINGDAL, M., TORSVIK, J., IMMERVOLL, H., HOEM, D., LAEMMERHIRT, F., SIMON, P., LERCH, M. M., JOHANSSON, S., NJØLSTAD, P. R., WEISS, F. U., FJELD, K. & MOLVEN, A. 2017. Copy number variants and VNTR length polymorphisms of the carboxyl-ester lipase (CEL) gene as risk factors in pancreatic cancer. *Pancreatology,* 17**,** 83-88.

FJELD, K., BEER, S., JOHNSTONE, M., ZIMMER, C., MÖSSNER, J., RUFFERT, C., KREHAN, M., ZAPF, C., NJØLSTAD, P. R., JOHANSSON, S., BUGERT, P., MIYAJIMA, F., LILOGLOU, T., BROWN, L., J. , WINN, S., A., DAVIES, K., LATAWIEC, D., GUNSON, B., K., CRIDDLE, D., N. , PIRMOHAMED, M., GRÜTZMANN, R., MICHL, P., GREENHALF, W., MOLVEN, A., SUTTON, R. & ROSENDAHL, J. 2016. Length of Variable Numbers of Tandem Repeats in the Carboxyl Ester Lipase (CEL) Gene May Confer Susceptibility to Alcoholic Liver Cirrhosis but Not Alcoholic Chronic Pancreatitis. *PLoS ONE,* 11**,** e0165567.

FJELD, K., WEISS, F. U., LASHER, D., ROSENDAHL, J., CHEN, J.-M., JOHANSSON, B., B., HOLGER, K., RUFFERT, C., MASSON, E., STEINE, S., J. , BUGERT, P., CNOP, M., GRÜTZMANN, R., MAYERLE, J., MÖSSNER, J., RINGDAL, M., SCHULZ, H.-U., SENDLER, M., SIMON, P., SZTROMWASSER, P., TORSVIK, J., SCHOLZ, M., TJORA,

E., FÉREC, C., WITT, H., LERCH, M., M. , NJØLSTAD, P., R., JOHANSSON, S. & MOLVEN, A. 2015. A recombined allele of the lipase gene CEL and its pseudogene CELP confers susceptibility to chronic pancreatitis. *Nature Genetics,* 47.
HAREL, T. & LUPSKI, J. R. 2017. Genomic disorders 20 years on—mechanisms for clinical manifestations. 93**,** 439-449.
HERNELL, O. & OLIVECRONA, T. 1974. Human milk lipases II. Bile salt-stimulated lipase. *Biochimica et Biophysica Acta,* 369**,** 234-244.
HOFSLI, E. 2018. *Bukspyttkjertelkreft* [Online]. Store medisinske leksikon. Available: https://sml.snl.no/bukspyttkjertelkreft [Accessed 21.05 2018].
HOLCK, P. 2017. *Bukspyttkjertelen* [Online]. Store medisinske leksikon Available: https://sml.snl.no/bukspyttkjertelen [Accessed 10.02.18 2018].
HOLMES, R. S. & COX, L. A. 2011. Comparative Structures and Evolution of Vertebrate Carboxyl Ester Lipase (CEL) Genes and Proteins with a Major Role in Reverse Cholesterol Transport. *Cholesterol,* 2011.
HUI, D. & HOWLES, P. 2002. Carboxyl ester lipase: structure-function relationship and physiological role in lipoprotein metabolism and atherosclerosis. *J. Lipid Res.*, 43**,** 2017-2030.
IN'T VELD, P. & MARICHAL, M. 2010. Microscopic Anatomy of the Human Islet of Langerhans. *Adv.Exp.Med.Biol.,* 654**,** 1-19.
JOHANSSON, B. B., FJELD, K., EL JELLAS, K., GRAVDAL, A., DALVA, M., TJORA, E., RÆDER, H., KULKARNI, R. N., JOHANSSON, S., NJØLSTAD, P. R. & MOLVEN, A. 2018. The role of the carboxyl ester lipase (CEL) gene in pancreatic disease. *Pancreatology,* 18**,** 12-19.
JOHANSSON, B. B., TORSVIK, J., BJØRKHAUG, L., VESTERHUS, M., RAGVIN, A., TJORA, E., FJELD, K., HOEM, D., JOHANSSON, S., RÆDER, H., LINDQUIST, S., HERNELL, O., CNOP, M., SARASTE, J., FLATMARK, T., MOLVEN, A. & NJØLSTAD, P. R. 2011. 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. *The Journal of biological chemistry,* 286**,** 34593-34605.
JUNQUEIRA, L. C. & CARNEIRO, J. 2003. *Basic Histology,* United States of America, McGraw-Hill Companies.
KAMISAWA, T., WOOD, L. D., ITOI, T. & TAKAORI, K. 2016. Pancreatic cancer. *The Lancet,* 388**,** 73-85.
KHARROUBI, A. T. & DARWISH, H. M. 2015. Diabetes mellitus: The epidemic of the century. *World journal of diabetes,* 6**,** 850-867.
KLEEFF, J., WHITCOMB, C. D., SHIMOSEGAWA, T., ESPOSITO, I., AKISIK, F., MUNOZ, D., ENRIQUE. J & NEOPTOLEMOS, P. J. 2017. Chronic pancreatitis. *Nature reviews* 3.
KUMAR, ABBAS & ESTER 2013. *Robbins Basic Pathology,* Canada, Elsevier saunders.
LEUNG, P. S. & IP, S. P. 2006. Pancreatic acinar cell: Its role in acute pancreatitis. *International Journal of Biochemistry and Cell Biology,* 38**,** 1024-1030.
LIDBERG, U., NILSSON, J., STRØMBERG, K., STENMAN, G., SAHLIN, P., ENERBÆCK, S. & BJURSELL, G. 1992. Genomic organization, sequence analysis, and chromosomal localization of the human carboxyl ester lipase (CEL) gene and a CEL-like (CELL) gene. *Genomics,* 13**,** 630-640.
LOMBARDO, D. 2001. Bile salt-dependent lipase: its pathophysiological implications. *BBA - Molecular and Cell Biology of Lipids,* 1533**,** 1-28.
LOMBARDO, D. & GUY, O. 1980. Studies on the substrate specificity of a carboxyl ester hydrolase from human pancreatic jucie. II. Action on cholesterol esters and lipid-soluble vitamin esters. *Biochimica et Biophysica Acta,* 611**,** 147-155.
LOMBARDO, D., GUY, O. & FIGARELLA, C. 1978. Purification and characterization of a carboxyl ester hydrolase from human pancreatic jucie. *Biochimica et Biophysica Acta,* 527**,** 142-149.
LOOMES, K. M., SENIOR, H. E., WEST, P. M. & ROBERTON, A. M. 1999. Functional protective role for mucin glycosylated repetitive domains. *European Journal of Biochemistry,* 266**,** 105-111.
MADEYSKI, K., LIDBERG, U., BJURSELL, G. & NILSSON, J. 1998. Structure and organization of the human carboxyl ester lipase locus. *Incorporating Mouse Genome,* 9**,** 334-338.

MARTINEZ, E., CRENON, I., SILVY, F., DEL GRANDE, J., MOUGEL, A., BAREA, D., FINA, F., BERNARD, J.-P., OUAISSI, M., LOMBARDO, D. & MAS, E. 2016. Expression of truncated bile salt-dependent lipase variant in pancreatic pre-neoplastic lesions. *Oncotarget,* 8**,** 536-551.

MARTINEZ, E., SILVY, F., FINA, F., BARTOLI, M., KRAHN, M., BARLESI, F., FIGARELLA-BRANGER, D., IOVANNA, J., LAUGIER, R., OUAISSI, M., LOMBARDO, D. & MAS, E. 2015. Rs488087 single nucleotide polymorphism as predictive risk factor for pancreatic cancers. *Oncotarget,* 6**,** 39855-39864.

MIYASAKA, K., OHTA, M., TAKANO, S., HAYASHI, H., HIGUCHI, S., MARUYAMA, K., TANDO, Y., NAKAMURA, T., TAKATA, Y. & FUNAKOSHI, A. 2005. Carboxylester Lipase Gene Polymorphism as a Risk of Alcohol-induced Pancreatitis. *Pancreas,* 30**,** e87-e91.

MOLVEN, A., FJELD, K. & LOWE, M. E. 2016. Lipase Genetic Variants in Chronic Pancreatitis: When the End Is Wrong, All's Not Well. *Gastroenterology,* 150**,** 1515-1518.

MOLVEN, A. & NJØLSTAD, P. R. 2011. Role of molecular genetics in transforming diagnosis of diabetes mellitus. *Expert Review of Molecular Diagnostics,* 11**,** 313-320.

NILSSON, J., HELLQUIST, M. & BJURSELL, G. 1993. The Human Carboxyl Ester Lipase-like (CELL) Gene Is Ubiquitously Expressed and Contains a Hypervariable Region. *Genomics,* 17**,** 416-422.

PANDOL, S. J. 2011. The exocrine pancreas. San Rafael (CA): Morgan & Claypool Life Sciences.

PASQUALINI, E., CAILLOL, N., VALETTE, A., LLOUBES, R., VERINE, A. & LOMBARDO, D. 2000. Phosphorylation of the rat pancreatic bile-salt-dependent lipase by casein kinase II is essential for secretion. *Biochem. J.,* 345**,** 121-128.

RAGVIN, A., FJELD, K., WEISS, F. U., TORSVIK, J., AGHDASSI, A., MAYERLE, J., SIMON, P., NJØLSTAD, P. R., LERCH, M. M., JOHANSSON, S. & MOLVEN, A. 2013. The number of tandem repeats in the carboxyl-ester lipase (CEL) gene as a risk factor in alcoholic and idiopathic chronic pancreatitis. *Pancreatology,* 13**,** 29-32.

REBOURS, V., LEVY, P. & RUSZNIEWSKI, P. 2012. An overview of hereditary pancreatitis. *Dig. Liver Dis.,* 44**,** 8-15.

REUE, K., ZAMBAUX, J., WONG, H., LEE, G., LEETE, T. H., RONK, M., SHIVELY, J. E., STERNBY, B., BORGSTRÖM, B. & AMEIS, D. 1991. cDNA cloning of carboxyl ester lipase from human pancreas reveals a unique proline-rich repeat unit. *Journal of lipid research,* 32**,** 267-276.

ROGERS, S., WELLS, R. & RECHSTEINER, M. 1986. Amino Acid Sequences Common to Rapidly Degraded Proteins: The PEST Hypothesis. *Science,* 234**,** 364-368.

RÆDER, H., JOHANSSON, S., HOLM, P., I. , HALDORSEN, I., S, MAS, E., SBARRA, V., NERMOEN, I., EIDE, S. Å., GREVLE, L., BJØRKHAUG, L., SAGEN, J., V., AKSNES, L., SØVIK, O., LOMBARDO, D., MOLVEN, A. & NJØLSTAD, P. R. 2006. Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction. *Nature Genetics,* 38**,** 54.

RÆDER, H., MCALLISTER, F. E., TJORA, E., BHATT, S., HALDORSEN, I., HU, J., WILLEMS, S. M., VESTERHUS, M., EL OUAAMARI, A., LIU, M., RÆDER, M. B., IMMERVOLL, H., HOEM, D., DIMCEVSKI, G., NJØLSTAD, P. R., MOLVEN, A., GYGI, S. P. & KULKARNI, R. N. 2014. Carboxyl-ester lipase maturity-onset diabetes of the young is associated with development of pancreatic cysts and upregulated MAPK signaling in secretin-stimulated duodenal fluid. *Diabetes,* 63**,** 259-269.

SAND, O., SJAASTAD, Ø. V., HAUG, E. & BJÅLIE, J. G. 2011. *Menneskekroppen*, Gyldendal Akademisk.

STAX, M. J., KOOTSTRA, N. A., VAN 'T WOUT, A. B., TANCK, M. W. T., BAKKER, M., POLLAKIS, G., PAXTON, W. A. & NIXON, D. F. 2012. HIV-1 Disease Progression Is Associated with Bile-Salt Stimulated Lipase (BSSL) Gene Polymorphism (HIV-1 Disease Links to BSSL Polymorphism). *PLoS ONE,* 7**,** e32534.

SØVIK, O., IRGENS, H. U., MOLNES, J., SAGENA, J. V., BJÖRKHAUG, L., RAEDER, H., MOLVENG, A. & NJÖLSTAD, P. R. 2013. Monogenic diabetes mellitus in Norway. *Monogenic diabetes mellitus in Norway*, 23.

TAMURA, K., YU, J., HATA, T., SUENAGA, M., SHINDO, K., ABE, T., MACGREGOR-DAS, A., BORGES, M., WOLFGANG, C. L., WEISS, M. J., HE, J., CANTO, M. I., PETERSEN, G.

M., GALLINGER, S., SYNGAL, S., BRAND, R. E., RUSTGI, A., OLSON, S. H., STOFFEL, E., COTE, M. L., ZOGOPOULOS, G., POTASH, J. B., GOES, F. S., MCCOMBIE, R. W., ZANDI, P. P., PIROOZNIA, M., KRAMER, M., PARLA, J., ESHLEMAN, J. R., ROBERTS, N. J., HRUBAN, R. H., KLEIN, A. P. & GOGGINS, M. 2018. Mutations in the pancreatic secretory enzymes CPA1 and CPB1 are associated with pancreatic cancer. *Proc Natl Acad Sci U S A,* 115**,** 4767-4772.

TAYLOR, A. K., ZAMBAUX, J. L., KLISAK, I., MOHANDAS, T., SPARKES, R. S., SCHOTZ, M. C. & LUSIS, A. J. 1991. Carboxyl ester lipase: A highly polymorphic locus on human chromosome 9qter. *Genomics,* 10**,** 425-431.

TERZYAN, S., WANG, C.-S., DOWNS, D., HUNTER, B. & ZHANG, X. C. 2000. Crystal structure of the catalytic domain of human bile salt activated lipase. *Protein Sci.,* 9**,** 1783-1790.

TORSVIK, J., JOHANSSON, B. B., DALVA, M., MARIE, M., FJELD, K., JOHANSSON, S., BJØRKØY, G., SARASTE, J., NJØLSTAD, P. R. & MOLVEN, A. 2014. Endocytosis of secreted carboxyl ester lipase in a syndrome of diabetes and pancreatic exocrine dysfunction. *The Journal of biological chemistry,* 289**,** 29097.

TORSVIK, J., JOHANSSON, S., JOHANSEN, A., EK, J., MINTON, J., RÆDER, H., ELLARD, S., HATTERSLEY, A., PEDERSEN, O., HANSEN, T., MOLVEN, A. & NJØLSTAD, P. 2010. Mutations in the VNTR of the carboxyl-ester lipase gene (CEL) are a rare cause of monogenic diabetes. *Human Genetics,* 127**,** 55-64.

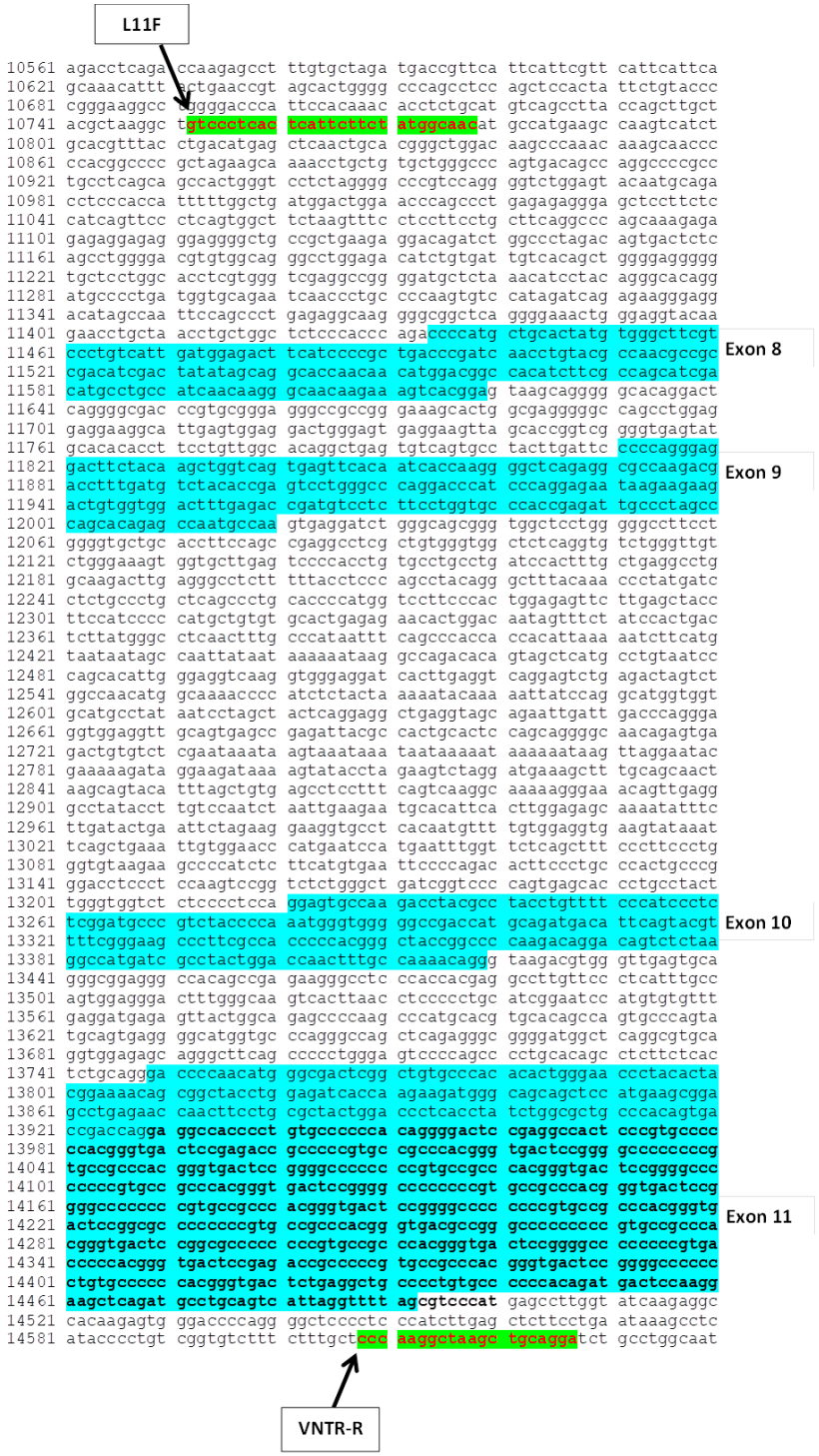
VANPUTTE, C., REGAN, J., RUSSO, A. & SEELY, R. 2014. *Seeley`s anatomy and Physiology* Mc Graw Hill.

XIAO, X., JONES, G., SEVILLA, W. A., STOLZ, D. B., MAGEE, K. E., HAUGHNEY, M., MUKHERJEE, A., WANG, Y. & LOWE, M. E. 2016. A carboxyl ester lipase (CEL) mutant causes chronic pancreatitis by forming intracellular aggregates that activate apoptosis. *The Journal of biological chemistry,* 292**,** 7744.

ØYRI, A. 2011. *Norsk medisinsk ordbok,* Oslo, Samlaget.

Appendix

Appendix 1. Binding sites for primers used in the Long-range PCR to amplify *CEL* exon 8-11. L11F is the forward primer (marked in green) and VNTR-R is the reverse primer (marked in green). The VNTR region in exon 11 is marked in bold letters.



Appendix 2. Binding sites for primers used to sequence *CEL* exon 11. EF is the forward primer (marked in green) and VNTR-R is the reverse primer (marked in green). The VNTR region in exon 11 is marked in bold letters.