Human Molecular Genetics, 2024, Vol. 33, 11, 1001–1014
https://doi.org/10.1093/hmg/ddae034
Advance access publication date 14 March 2024
Original Article

Common single-base insertions in the VNTR of the carboxyl ester lipase (*CEL*) gene are benign and also likely to arise somatically in the exocrine pancreas

Ranveig S. Brekke[1,2,3], Anny Gravdal[1,2,3], Khadija El Jellas[1,2], Grace E. Curry[4], Jianguo Lin[4], Steven J. Wilhelm[4], Solrun J. Steine[1], Eric Mas[5], Stefan Johansson[2,3], Mark E. Lowe[4], Bente B. Johansson[2], Xunjun Xiao[4], Karianne Fjeld[1,2,3], Anders Molven[1,2,6,*]

[1]Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen, Jonas Lies vei 91B, 5021 Bergen, Norway
[2]Center for Diabetes Research, Department of Clinical Science, University of Bergen, Jonas Lies vei 87, 5021 Bergen, Norway
[3]Department of Medical Genetics, Haukeland University Hospital, Jonas Lies vei 91B, 5021 Bergen, Norway
[4]Department of Pediatrics, Washington University School of Medicine, Campus Box 8208, 660 South Euclid Ave, St. Louis, MO 63110, USA
[5]Cancer Research Center of Marseille, Aix Marseille University, CNRS, INSERM, Institut Paoli-Calmettes, CRCM, 27 Bd Leï Roure, 13273 Marseille Cedex 09, France
[6]Department of Pathology and Section for Cancer Genomics, Haukeland University Hospital, Jonas Lies vei 83, Bergen, Norway

*Corresponding author. Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen, Jonas Lies vei 91B, Bergen 5021, Norway.
E-mail: anders.molven@uib.no

Abstract

The *CEL* gene encodes carboxyl ester lipase, a pancreatic digestive enzyme. *CEL* is extremely polymorphic due to a variable number tandem repeat (VNTR) located in the last exon. Single-base deletions within this VNTR cause the inherited disorder MODY8, whereas little is known about VNTR single-base insertions in pancreatic disease. We therefore mapped *CEL* insertion variants (*CEL-INS*) in 200 Norwegian patients with pancreatic neoplastic disorders. Twenty-eight samples (14.0%) carried *CEL-INS* alleles. Most common were insertions in repeat 9 (9.5%), which always associated with a VNTR length of 13 repeats. The combined *INS* allele frequency (0.078) was similar to that observed in a control material of 416 subjects (0.075). We performed functional testing in HEK293T cells of a set of *CEL-INS* variants, in which the insertion site varied from the first to the 12th VNTR repeat. Lipase activity showed little difference among the variants. However, *CEL-INS* variants with insertions occurring in the most proximal repeats led to protein aggregation and endoplasmic reticulum stress, which upregulated the unfolded protein response. Moreover, by using a CEL-INS-specific antibody, we observed patchy signals in pancreatic tissue from humans without any *CEL-INS* variant in the germline. Similar pancreatic staining was seen in knock-in mice expressing the most common human *CEL* VNTR with 16 repeats. CEL-INS proteins may therefore be constantly produced from somatic events in the normal pancreatic parenchyma. This observation along with the high population frequency of *CEL-INS* alleles strongly suggests that these variants are benign, with a possible exception for insertions in VNTR repeats 1–4.

Keywords: carboxyl ester lipase; pancreatic cancer; single-base insertions; variable number tandem repeat

Introduction

Carboxyl ester lipase (CEL) is expressed in the acinar cells of the pancreas and is one of four major lipases secreted by the exocrine pancreas [1]. CEL is also expressed in lactating mammary glands and secreted with the mother's milk [2]. The enzyme hydrolyzes cholesterol esters, dietary fat and fat-soluble vitamins, and bile salts stimulate the lipase activity [3, 4]. Hence, CEL is also referred to as bile salt-stimulated lipase (BSSL) [2] or bile salt-dependent lipase (BSDL) [5].

The *CEL* gene covers approximately 10 kilo-base pairs (bp) and contains 11 exons, of which the last exon includes a variable number tandem repeat (VNTR) [6]. The length of each VNTR repeat is 33 bp, encoding a protein segment of 11 amino acids. The number of VNTR repeats vary from three to 23 with 16 repeats being the most common [7–10]. In addition to VNTR length variation, single-base insertions and deletions inside the VNTR region and copy number variants involving the *CEL* locus have been described, implying that *CEL* is among the most polymorphic genes in the human genome [6, 11, 12].

Disease risks associated with the *CEL* gene are mainly attributed to variants of the VNTR. Thus, VNTR length polymorphisms have been linked to alcoholic liver cirrhosis and chronic pancreatitis [10, 13]. In addition, a copy number variant of *CEL* (*CEL-HYB1*), in which the VNTR stems from a *CEL* pseudogene, is a well-established risk factor for chronic pancreatitis [14–16]. The strongest disease association, however, is seen for maturity-onset diabetes of the young, type 8 (MODY8, also known as CEL-MODY). This inherited disorder was first reported in 2006 in two Norwegian families that suffered from combined pancreatic exocrine dysfunction and diabetes [17]. Since then, three additional families have been published [18, 19]. Hallmarks of MODY8 are low fecal elastase levels, pancreatic fatty infiltration

Received: November 24, 2023. **Revised:** February 14, 2024. **Accepted:** February 27, 2024

and insulin-dependent diabetes, eventually arising in all mutation carriers [19, 20]. The causative mutations are single-base deletions within one of the first five repeats of the *CEL* VNTR, resulting in a somewhat shorter CEL protein with a completely changed C-terminal due to an altered reading frame and a premature stop codon [19]. Cellular studies indicate that the disease mechanism involves protein misfolding and aggregation, followed by cellular stress, cytotoxicity, and cell death [21–24].

The study identifying MODY8 also noted that some *CEL* alleles contain single-base insertions in the VNTR sequence [17]. Like single-base deletions, these insertions result in a premature stop codon and truncated CEL protein variants [17]. However, for almost all insertions, the new stop codon is predicted to arise within the same repeat in which the mutation occurred. The insertion of an extra nucleotide in the VNTR will in most cases change the predicted C-terminal amino-acid sequence from PAVIRF in normal CEL into the sequence PRAAHG (Fig. 1). Martinez *et al*. have developed antibodies directed towards either the PAVIRF or the PRAAHG terminus of the CEL protein and suggested that single-base insertions of the *CEL* VNTR occur early during pancreatic carcinogenesis [25].

CEL is now included in diagnostic gene panels for both monogenic diabetes [26] and chronic pancreatitis [27]. As all pathogenic variants of *CEL* known so far change the VNTR region, it is necessary to assess prevalence and properties of the single-base insertions within the *CEL* VNTR, which we refer to as *CEL-INS*. We therefore performed genetic screening for *CEL-INS* alleles in human samples, tested the variants functionally, and analyzed pancreatic tissue from humans and mice for CEL-INS protein expression.

Results

The prevalence of single-base insertion variants in the *CEL* VNTR

We first analyzed 200 DNA samples from a Norwegian cohort of patients with neoplastic disorders of the pancreas [28, 29]. The screening method consisted of a long PCR to avoid amplification of the *CEL* pseudogene sequence [14], followed by Sanger sequencing of the *CEL* VNTR from both the 5′ and 3′ end. The *CEL* VNTR is very GC-rich and contains many stretches of homopolymeric tracts, resulting in electropherograms that need to be interpreted manually when amplifying and sequencing from genomic DNA (Fig. 2A). This is particularly important for those VNTR segments that are positioned distally from the start of the sequencing reaction.

We detected 28 cases with single-base insertions in the *CEL* VNTR among the 200 screened samples, including three cases that were homozygous. This corresponded to a carrier frequency of 0.14 and an allele frequency of 0.078 (Table 1). *INS* alleles were more frequently observed in males (n = 17, 17.9%) than in females (n = 11, 10.5%), but the gender difference was not statistically significant ($P = 0.13$). Twenty *INS*-positives were found among 159 cases classified as malignant lesions, whereas the remaining eight *INS*-positives were identified among the 41 cases with a benign diagnosis (difference not statistically significant, $P = 0.25$).

Within the patient material, the observed *INS* variants were positioned in repeat 9 (*INS9*, n = 19), repeat 10 (*INS10*, n = 2) and repeat 12 (*INS12*, n = 7) (Table 1). Notably, all subjects that tested positive for *INS9*, the most frequent insertion variant, carried a *CEL* VNTR consisting of 13 repeated segments. This included three *INS9* homozygous samples that were homozygous also for the 13-repeat VNTR. Hence, in our material, the *INS9* variant was always present on a *CEL* allele with 13 VNTR repeats. However, the association was not absolute in the other direction: Among the 200 screened patients, there were 25 samples positive for a 13-repeat VNTR and six (24.0%) of these did not carry any single-base insertion. *INS10* and *INS12* were associated with 14- and 16-repeat VNTRs, respectively (Table 1).

To investigate whether there could be an association between *CEL-INS* variants and pancreatic cancer, we compared the carrier frequencies above with frequencies observed in a population-based control material (n = 416, see Materials and Methods). Notably, some additional types of *INS* variants (*INS4*, *INS8*, *INS11*) were observed among the controls (Supplementary Table 1). The overall *INS* carrier frequency in the patient material was almost identical to that of the controls (0.140 vs. 0.139; OR = 1.00, 95% CI = 0.62–1.63, $P = 1.0$). We also tested the carrier frequences of the two most common *INS* variants. There were no statistically significant differences when cases and controls were compared (*INS9*: OR = 0.96, 95% CI = 0.54–1.70, $P = 1.0$; *INS12*: OR = 1.64, 95% CI = 0.60–4.47, $P = 0.33$).

Only one common single-base substitution in the *CEL* VNTR

During the analysis of genomic DNA, we sometimes encountered single-base substitutions in the *CEL* VNTR. However, the only substitution commonly observed was a change from C to T in position 12 of the second VNTR segment (Fig. 2B). This is a single-nucleotide polymorphism (SNP) registered in the genome databases as variant rs488087. There were 105 subjects with the genotype C/C (52.5%), 76 with C/T (38.0%) and 19 with T/T (9.5%). These values corresponded to allele frequencies for C and T of 71.5% and 28.5%, respectively, which is the same frequencies as observed in the ExAC dataset [30]. Notably, all *INS9*-positive samples carried the minor T allele.

Enzymatic activity is minimally affected in the CEL-INS variants

Single-base insertions in the *CEL* VNTR change the reading frame of the protein. The predicted consequence is premature translation termination and truncated CEL proteins with a C-terminal sequence different from that of normal CEL (Fig. 1; Fig. 2C). To assess whether CEL's primary function could be affected, we measured lipase activity for our set of INS variants. Conditioned medium was collected from transfected HEK293T cells, and each protein variant was purified to near homogeneity (Fig. 3A). Enzymatic activity in the medium was quantified with trioctanoin as substrate. The measured activities (Fig. 3B) were normalized against the molecular weight of the variants (Fig. 1) and adjusted for relative band intensities (Fig. 3A) to compensate for variations in the number of molecules. Although a statistically significant increase was observed for CEL-INS12, all insertion variants showed enzyme activities very comparable to that of CEL-WT, i.e. the normal CEL protein (Fig. 3B).

Effects of CEL insertions variants on secretion and intracellular distribution

To determine if single-base insertions within the CEL VNTR influence secretion and subcellular distribution of the protein, we tested the effect of the insertions in transiently transfected HEK293T cells by employing the set of constructs depicted in Fig. 1. The conditioned medium, whole-cell lysate, soluble cell lysate fraction and the insoluble pellet fraction were analyzed by western blotting (Fig. 4). When secretion was assessed (Fig. 4A), the strongest effect was seen for CEL-INS1 (reduced by 76%

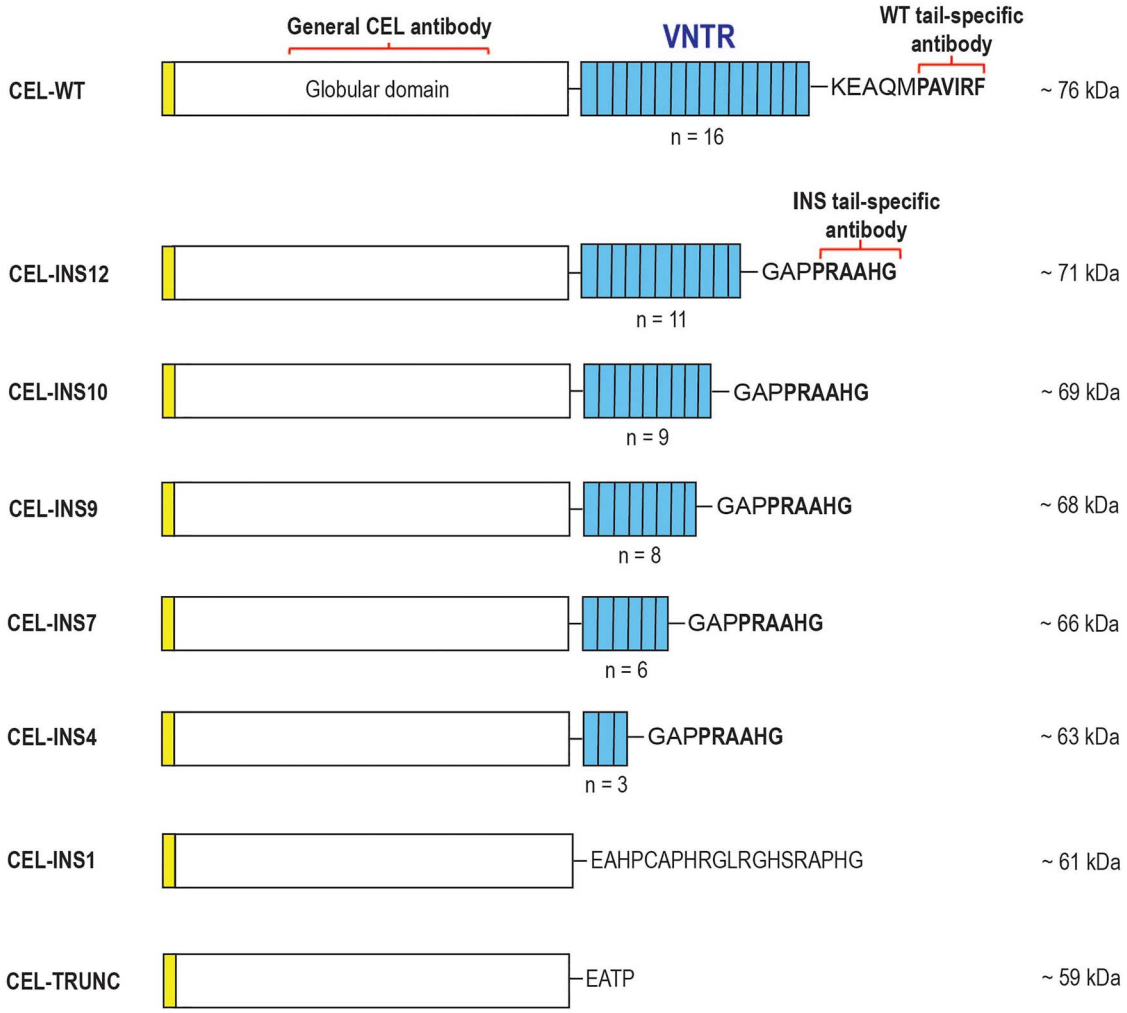

Figure 1. Schematic overview of the different CEL protein variants investigated. CEL-WT (wild-type) corresponds to the most common CEL variant in the general population and contains 16 VNTR repeats, each consisting of 11 amino acids. The general CEL antibody used in this study recognizes the globular domain of all variants, whereas the tail-specific antibody anti-PAVIRF targets the six terminal amino acids of CEL-WT. The tail-specific antibody anti-PRAAHG targets the six terminal amino acids of INS4-INS12. For these variants, a premature stop codon occurs within the same VNTR repeat affected by the insertion. Due to a difference in DNA sequence of VNTR repeat 1, translation of the INS1 construct will not terminate until repeat 2, giving rise to a C-terminal sequence differing from the other INS variants. CEL-TRUNC is truncated variant of CEL and contains only the first four amino acids (EATP) of the CEL-WT VNTR region. The yellow box to the left indicates the N-terminal signal peptide, whereas light blue boxes represent the number (n) of normal repeats encoded by the VNTR. Predicted molecular mass (kDa) for each CEL variant, without signal peptide and post-translational modifications, is shown on the right. Elements of the figure are not drawn to scale.

Table 1. Frequency of single-base insertion variants in the VNTR region of *CEL* in 200 patients with neoplastic pancreatic disease.

Variant	Number of positive samples	Allele frequency (n = 400)	Carrier frequency (n = 200)	Observed VNTR lengths
INS9	19[a]	0.055	0.095	13/13–17[b]
INS10	2	0.005	0.010	14/15, 16[c]
INS12	7	0.018	0.035	16/14, 16[d]
Total	**28**	**0.078**	**0.140**	-

[a]Among the positive samples, three were homozygous for the *INS9* allele. [b]The distribution of VNTR lengths for the *INS9*-positive samples was: 13/13 (n = 2), 13/14 (n = 2), 13/15 (n = 3), 13/16 (n = 10) and 13/17 (n = 2). [c]For *INS10*-positive samples: 14/15 (n = 1), 14/16 (n = 1). [d]For *INS12*-positive samples: 14/16 (n = 1), 16/16 (n = 6).

compared to CEL-WT). Likewise, secretion of the truncated control construct CEL-TRUNC (see Methods) was reduced by 48%. The other CEL-INS variants had secretions levels similar to CEL-WT (Fig. 4A). We also measured lipase activity in the conditioned medium from the transfected HEK293T cells. The varying enzyme activities of the CEL-INS variants (Supplementary Fig. 1) corresponded closely to the quantification of band intensities in Fig. 4A, supporting the conclusion that enzymatic activity differences between the INS variants and the normal CEL protein are very modest (Fig. 3).

In the whole-cell lysates, the most prominent difference was a significant increase in the abundance of CEL-INS1 by 127% and CEL-TRUNC by 94% compared to CEL-WT (Fig. 4B). As the position of the insertion became more distal, the abundance appeared to

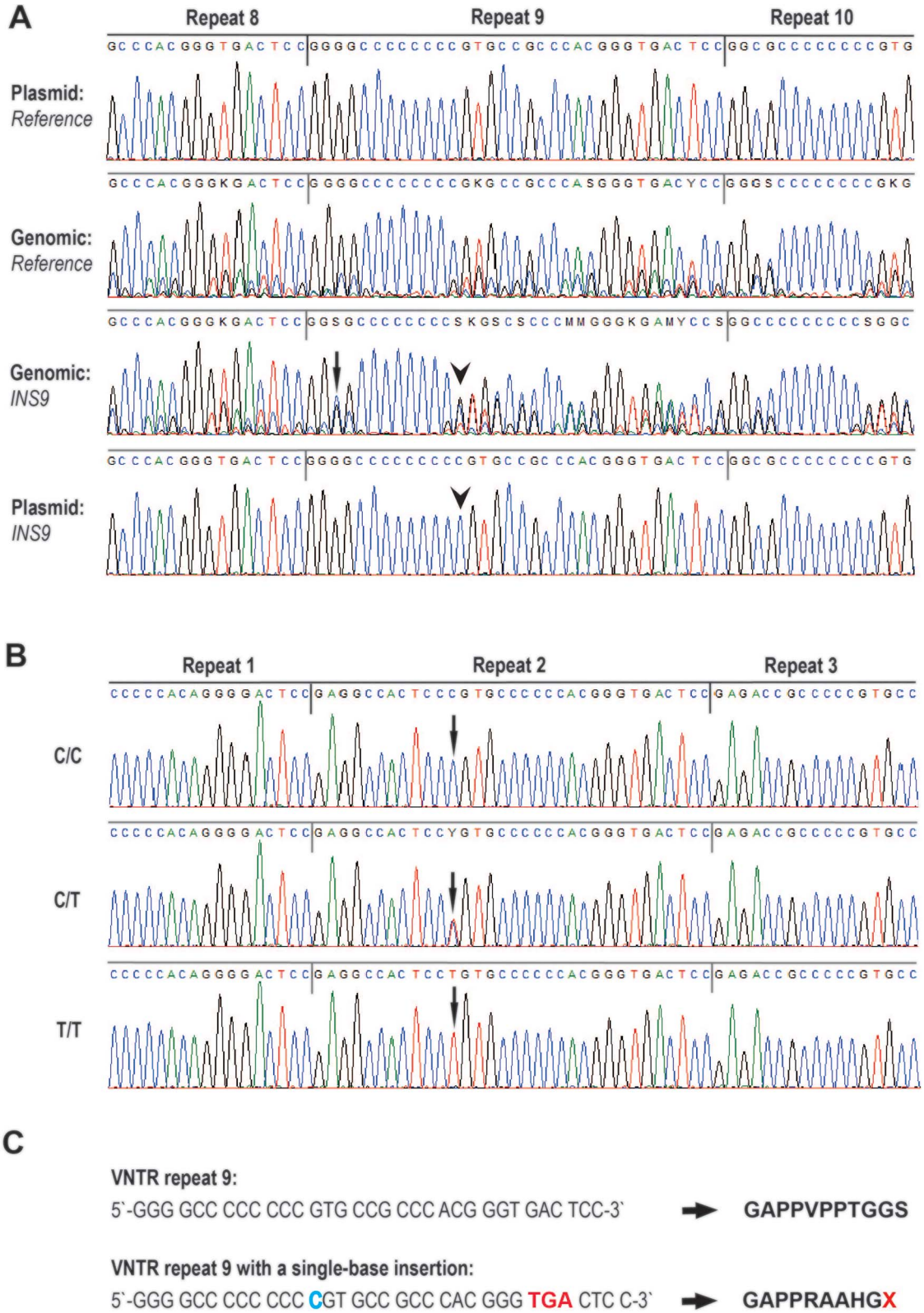

Figure 2. DNA sequence analysis of the *CEL* VNTR region. (A) Identification of single-base insertions within the VNTR. Electropherograms from sanger sequencing of plasmid constructs and human genomic DNA are compared. Note the lower technical quality obtained when sequencing the human samples. VNTR repeat number 9 and parts of the neighboring repeats 8 and 10 are shown. The two upper panels correspond to *CEL* reference sequences without any variants. The third sequence has been obtained from a human sample heterozygous for an extra C (arrowhead) inserted in the poly-C tract of VNTR repeat 9 (i.e. the variant designated *CEL-INS9*). The electropherogram starts to differ from the reference sequence at this position. In the same sample, a heterozygous single-base substitution (G > C) is present in position 3 of repeat 9 (arrow). The lower sequence is from the INS9 plasmid construct, with the arrowhead pointing to the inserted C. (B) Identification of the common single-base substitution (SNP rs488087, arrow) in position 12 of repeat 2 of the *CEL* VNTR. Electropherograms of human genomic DNA corresponding to the three different genotypes C/C, C/T and T/T are shown. Note the good technical quality when sequencing proximal VNTR repeats. Only parts of the neighboring repeats 1 and 3 are displayed. (C) Predicted effect of the *INS9* insertion variant on the CEL protein. The bases of *CEL* VNTR repeat 9 are shown with and without the inserted C (blue). The premature stop codon is highlighted in red. Predicted protein sequences are shown to the right.

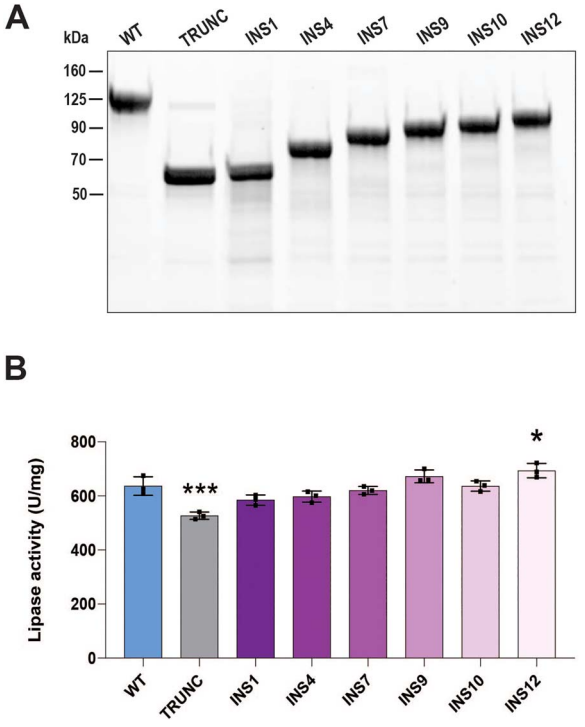

Figure 3. Lipase activity of CEL insertions variants. HEK293T cells were transiently transfected with the *CEL* constructs depicted in Fig. 1, and the recombinant CEL protein variants were purified to homogeneity from conditioned medium. (A) Purified CEL protein variants separated by SDS-PAGE and stained with GelCode Blue. An equal amount of protein was loaded in each lane. The image is representative for n = 3 independent experiments. (B) Lipolytic activity of purified CEL protein variants with trioctanoin as substrate. For each variant, activity was adjusted for band density (panel A) and molecular mass (Fig. 1). Error bars are SD. Statistical significance for activity different from that of the CEL-WT protein is marked with asterisks (* = $P \leq 0.05$; *** = $P \leq 0.001$).

gradually decrease. The pattern in the soluble lysate was opposite: The lowest levels were observed for CEL-TRUNC (83% reduction compared to CEL-WT), INS1 (70%), INS4 (74%) and INS7 (56%), with a general increase in abundance as the position of the insertion moved distally (Fig. 4C). In this fraction, all variants except INS10, differed significantly from CEL-WT. Finally, we assessed the insoluble pellet. Here, CEL-TRUNC and all insertion variants except INS12 were significantly more abundant than CEL-WT, with profound increases varying from 12- to 22-fold (Fig. 4D). Among the insertion variants, the effect was strongest for INS1, and it then decreased as the position of insertion became more distal.

Effects on endoplasmic reticulum (ER) stress markers

We previously demonstrated that single-base deletions within the *CEL* VNTR led to increased expression of ER stress markers [24]. It was therefore of interest to perform a similar analysis for the *CEL* insertion variants. HEK293T cells were transfected by our set of *INS* constructs (Fig. 1), and immunoblotting of whole-cell lysates and the insoluble pellet fraction was used to determine the level of the ER stress transducer PERK and the chaperones calnexin, BiP and calreticulin as markers for the unfolded protein response (UPR). In the whole-cell lysate, significantly higher levels of BiP were detected for CEL-INS1 (increased by 96% compared to CEL-WT and CEL-TRUNC (70% increase) (Fig. 5A). Moreover, somewhat lower levels of calnexin, in the range of 20–30% reduction, were observed for CEL-INS9, INS10 and INS12.

In the insoluble pellet, the strongest effects were observed for the variants CEL-INS1 and CEL-INS4. Calnexin, BiP and calreticulin were significantly increased for these two variants with levels at least 7-fold higher than when CEL-WT was expressed (Fig. 5B). In addition, expression of CEL-INS1 increased the level of PERK almost 4-fold. For CEL-INS variants 7–12, levels of the stress markers were comparable to that of CEL-WT. In line with the experiment presented in Fig. 4, the more distal the insertion mutation was positioned, the more similar to the CEL-WT response did the stress marker levels appear (Fig. 5B).

CEL-INS protein variants are expressed in human pancreatic tissue

Since some CEL-INS protein variants exhibit cellular properties that differ significantly from those of normal CEL, it was of interest to check whether INS variants are expressed in the human pancreas. A general anti-CEL antibody would, however, be of little information value as the great majority of *INS*-positive carriers are heterozygous and will express normal CEL protein from one allele. We therefore took advantage of antibodies directed towards the unique C-terminal amino acids in normal CEL (=anti-PAVIRF) and in the INS variants (=anti-PRAAHG) (Fig. 1). These antibodies were first tested by immunoblotting of cell lysates from transfected HEK293 cells. As predicted, the general antibody against the globular domain detected both CEL-WT, CEL-INS9 and CEL-TRUNC. Moreover, the anti-PAVIRF antibody detected only the normal, full-length CEL protein. The anti-PRAAHG antibody detected the INS9 variant but not the two other tested proteins (Supplementary Fig. 2A). Reactivity was confirmed by immunofluorescent staining and confocal imaging of the transfected cells (Supplementary Fig. 2B).

Tissue sections of human pancreas were then subjected to immunohistochemistry by using the antibodies tested above. Specimens from two *INS*9-positive patients were stained along with a control pancreatic tissue section from a patient who was negative for single-base insertions. The three specimens were from patients with pancreatic cancer, but we analyzed resected tissue that was located away from the tumor and had normal morphology. The two INS9 carriers showed a similar, positive staining pattern of the acinar cells for all three antibodies, confirming that CEL protein was expressed both from the *INS*9 and the normal *CEL* allele (Fig. 6A).

For the *INS*-negative control section, the staining pattern was the same when employing the antibody against the globular CEL-domain (anti-CEL) or against the WT-specific tail (anti- PAVIRF). However, quite unexpectedly, when staining the *INS*-negative control with the INS-specific anti-PRAAHG antibody, we observed areas of patchy, positive staining (Fig. 6A). Acinar-ductal metaplasia and PanIN precursor lesions close to the tumor area were consistently negative (data not shown).

To exclude any effect of malignant disease, we stained normal pancreatic tissue resected from four different patients with benign lesions and who were from 18 to 74 years old. To our surprise, all four samples showed clusters of pancreatic acini or single acinar cells with positive signal when stained with the INS-specific anti-PRAAHG antibody (Fig. 6B).

Detection of CEL-INS protein in mouse pancreatic tissue

Based on the positive anti-PRAAHG staining observed in pancreatic tissue obtained from subjects who were negative for *CEL-INS*

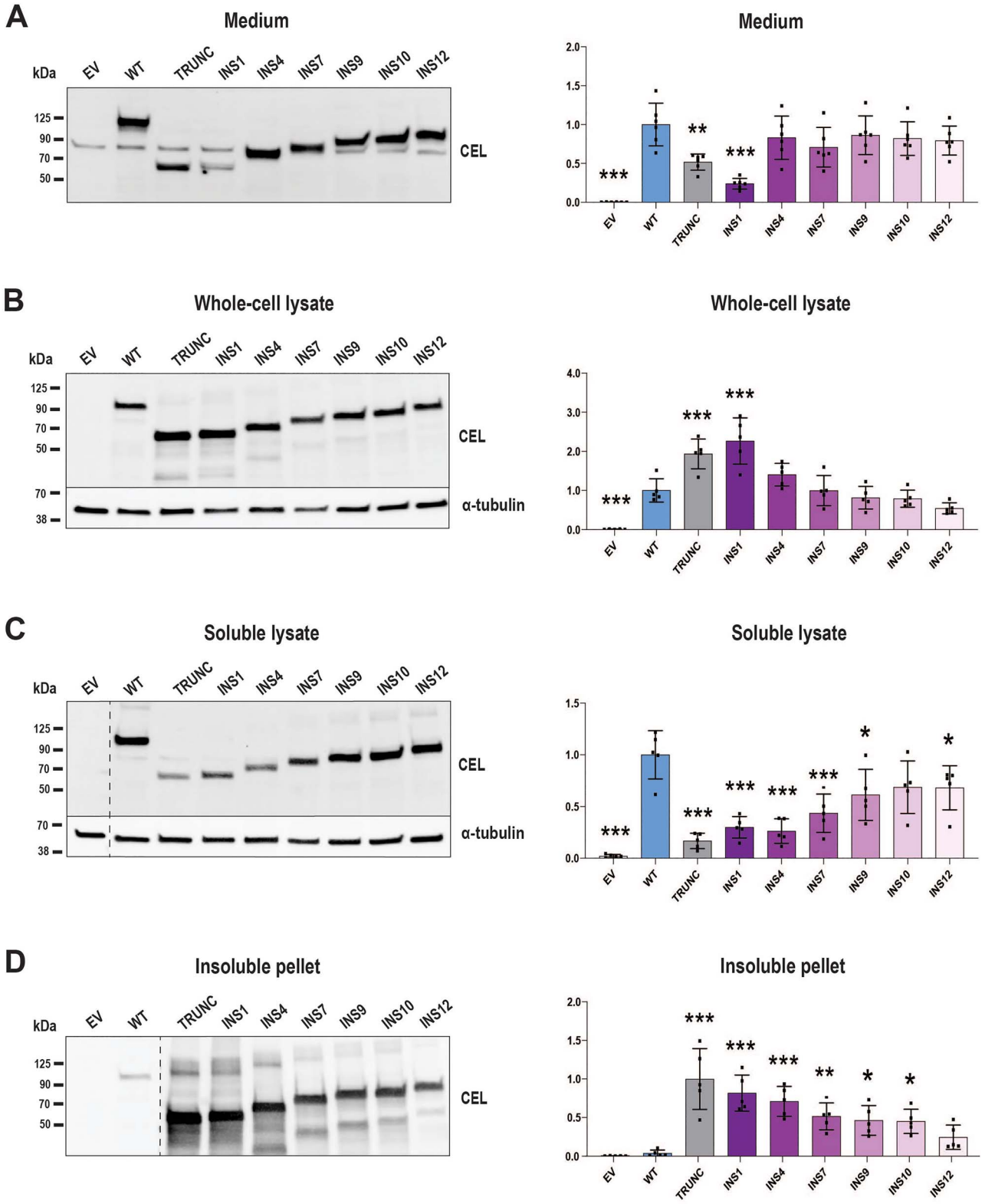

Figure 4. Secretion and cellular distribution of CEL insertion variants. HEK293T cells were transiently transfected with the *CEL* constructs depicted in Fig. 1. The western blots show the level of CEL protein in different cellular fractions. α-tubulin was used as loading control and empty vector (EV) served as negative control. The quantitative data to the right are normalized to the abundance of CEL-WT, except for the insoluble pellet where CEL-TRUNC was used for normalization as the very low abundance of CEL-WT is not a reliable denominator. Band intensities of the lysates were first adjusted to the α-tubulin level. (A) Secretion of CEL variants measured in the conditioned medium. An unspecific band around 85 kDa is observed for all constructs. (B–D) Intracellular distribution assessed after lysis of the cells by analyzing the whole-cell lysate (B), the soluble lysate (C) and the insoluble pellet (D). For each fraction, representative images from n = 6 independent experiments are shown. Dashed lines indicate an unrelated lane removed from the images. Numbers on the Y-axis represent relative protein amounts. Error bars are SD. Statistical significance for band intensities different from the CEL-WT intensity is marked with asterisks (* = $P \leq 0.05$; ** = $P \leq 0.01$; *** = $P \leq 0.001$).

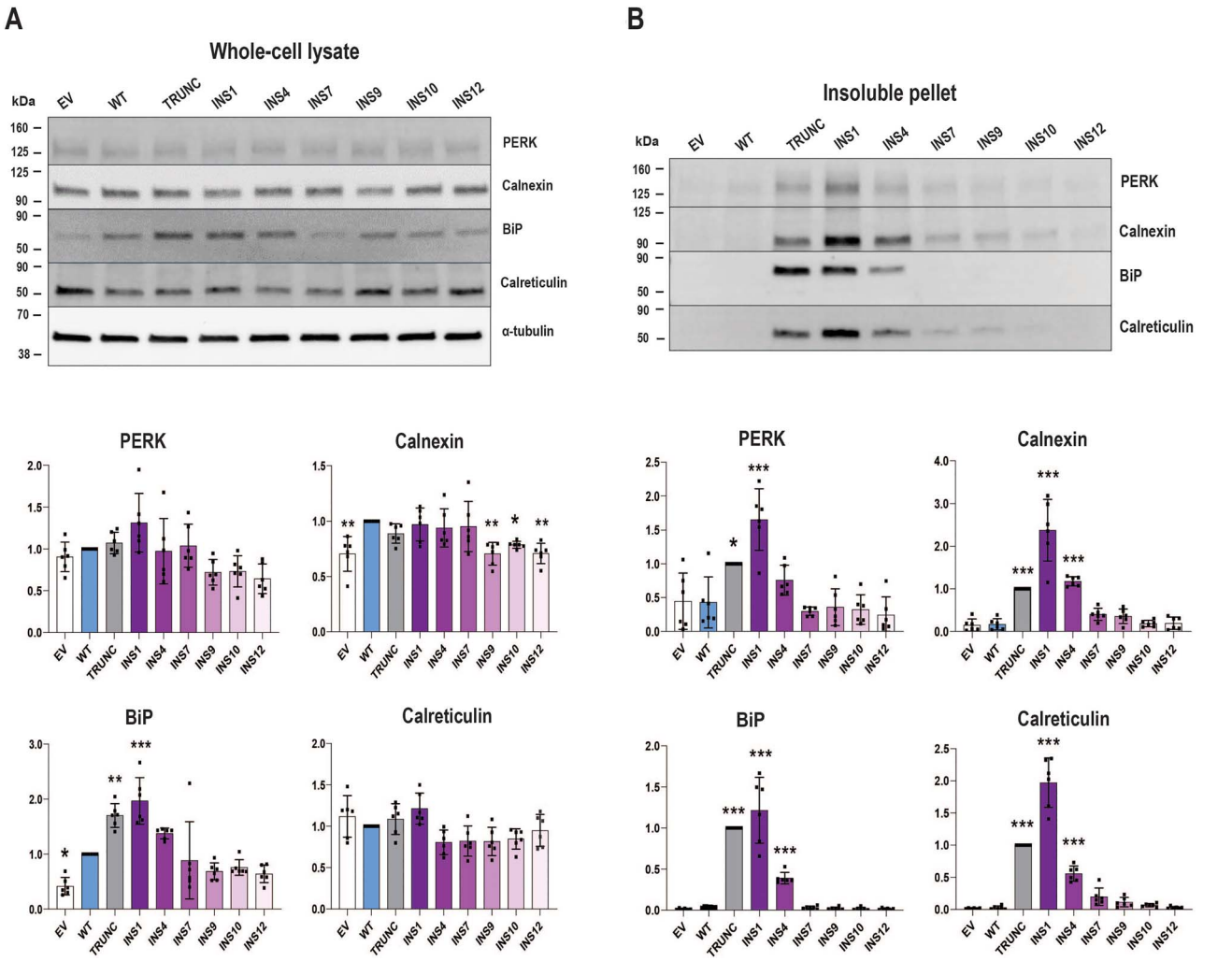

Figure 5. Effect of CEL insertion variants on ER stress and UPR markers. HEK293T cells were transiently transfected with the *CEL* constructs depicted in Fig. 1. Cells were harvested 24 h after medium change, i.e. 68–72 h after transfection. (A) Western blot showing levels of ER stress and UPR markers in the whole-cell lysate. α-tubulin was used as loading control and empty vector (EV) served as negative control. Representative images from n = 6 independent experiments are shown for each of PERK, calnexin, BiP (GRP78) and calreticulin. Quantification of band intensities is shown below the blot, adjusted for α-tubulin level and then normalized to the marker level for the CEL-WT construct. Numbers on the Y-axis represent relative protein amounts. Error bars are SD. Statistical significance for band intensities different from that of the CEL-WT experiment is marked with asterisks (* = $P \leq 0.05$; ** = $P \leq 0.01$; *** = $P \leq 0.001$). (B) Data for the insoluble pellet fraction obtained from the same experiment as in (A). Legend is the same, except that band intensities were normalized to the marker level for the CEL-TRUNC construct due to very low marker levels when transfecting with the CEL-WT construct.

variants at the germline level, we postulated that such variants might be formed by somatic, mutational events. To test this hypothesis, we employed knock-in mice in which the normal VNTR of the mouse *Cel* gene (consisting of only three repeats) had been substituted with the human *CEL* VNTR consisting of 16 repeats (Fig. 7A; Supplementary Fig. 3). This mouse strain is designated *Cel-16R*, and its pancreatic morphology is undistinguishable from the pancreas of normal mice (Supplementary Fig. 4). We performed immunohistochemistry with the anti-PRAAHG antibody and examined pancreatic tissues from mice aged 3, 6 and 12 months. Sections were obtained from heterozygous *Cel-16R* knock-in mice and C57Bl/6 N wild-type animals (Fig. 7B). Specimens from wild-type mice were consistently negative whereas patchy, positive staining reminiscent of the patterns in Fig. 6B were observed in the pancreatic sections from the knock-in mice. We counted the number of INS-positive cells, and the average proportion was 0.5%, 1.5% and 1.2% of all cells in the tissue for mice aged 3, 6 and 12 months, respectively (Fig. 7B). Although there was a trend of higher proportions in the older mice, the difference between groups was not statistically significant. We also stained homozygous *Cel-16R* mice (6 and 12 months). The proportion of positive cells was similar to that of heterozygous animals (data not shown).

Discussion

CEL mutations can either cause or increase the risk for pancreatic disease. So far, all *CEL* variants that are known to be pathogenic, involve alterations of the gene's VNTR sequence [6]. The strongest association has been demonstrated for the very rare, single-base deletions within the VNTR that cause MODY8 [17–19]. In addition, *CEL-HYB1*, a hybrid allele of *CEL* and the VNTR region of the nearby pseudogene *CELP*, is a risk factor for developing chronic pancreatitis [14]. *CEL* is now part of diagnostic gene panels for inherited diabetes and pancreatitis [26, 27] but, unfortunately, an array of papers have in recent years reported proposed disease-causing *CEL* variants without any presentation of functional evidence or co-segregation in pedigrees (see the Discussion in [19]).

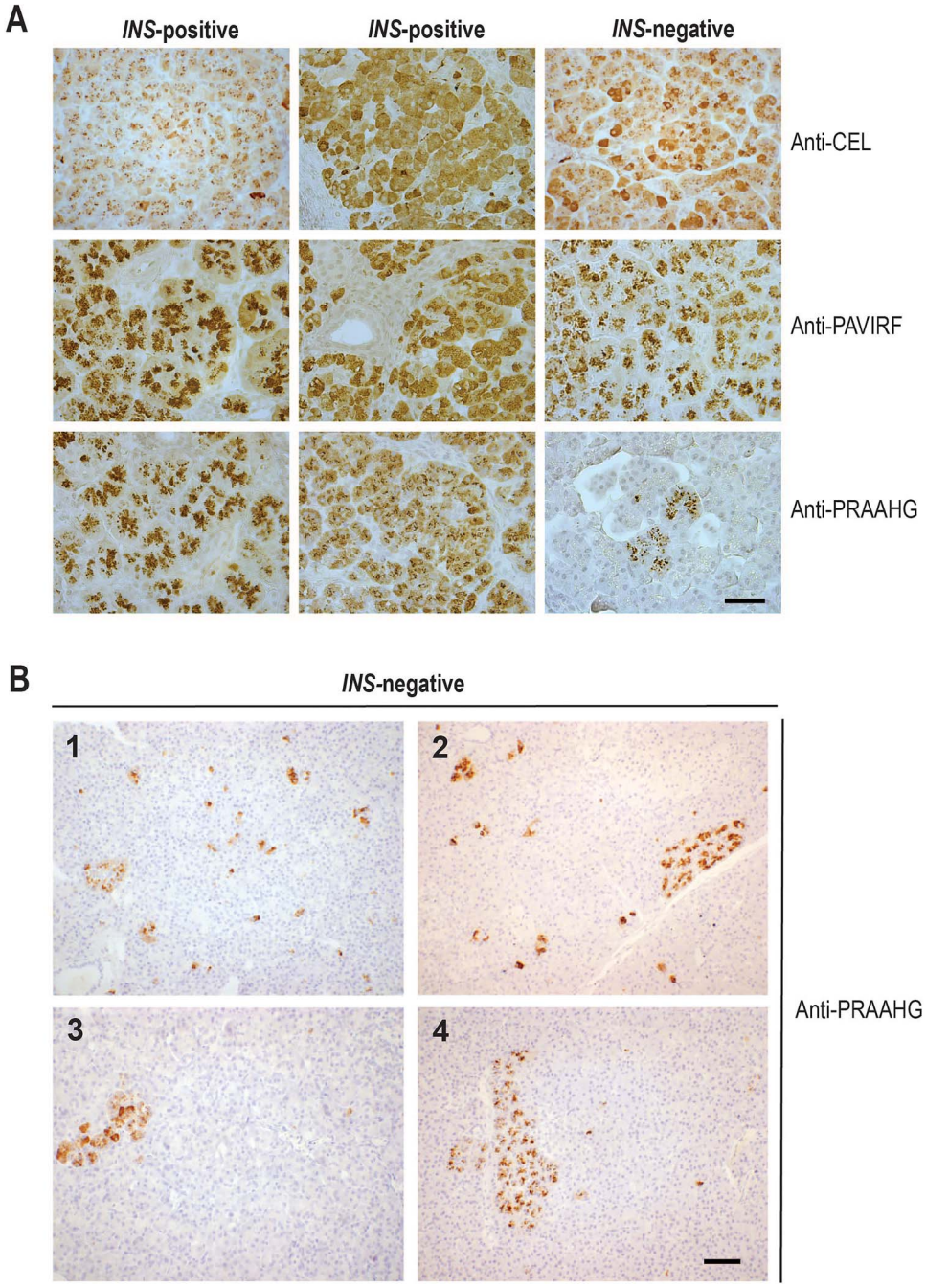

Figure 6. Reactivity of a CEL-INS-specific antibody in human pancreatic tissue. (A) Staining of morphologically normal tissue sections from three pancreatic cancer patients. Two cases were *CEL-INS*9-positive, and one case had no single-base insertion in the VNTR (=*INS*-negative). The sections were stained with the antibodies anti-CEL, anti-PAVIRF and anti-PRAAHG (see Fig. 1). Positive CEL staining is shown in dark brown color. Scale bar = 40 μm. (B) Reactivity of the anti-PRAAHG antibody in pancreatic tissue obtained from four patients (1–4) without a malignant diagnosis and not carrying germline *INS* variants. Scale bar = 80 μm.

In contrast to the pathogenic single-base deletions, there is little knowledge about the impact of the more commonly observed single-base insertions within the *CEL* VNTR. To the best of our knowledge, *CEL-INS* variants have only been studied in two papers of the literature. Ræder *et al.* [17] investigated a material of 182 persons with diabetes and did suggest an impact of insertion variants on pancreatic function as these variants were overrepresented among a subset of 20 cases with low fecal elastase levels. Martinez *et al.* [25] examined *CEL* VNTR single-base insertions in the context of pancreatic adenocarcinoma (32 cases) and proposed that such variants could be involved in early pancreatic carcinogenesis. The aim of our study was therefore to present an in-depth assessment of the pathogenic potential of *CEL-INS* variants to facilitate correct interpretation of them in the clinical setting.

CEL-INS variants are notoriously difficult to mine from high-throughput data sets produced by next-generation sequencing technology. This is due to the high GC content and repetitive nature of the *CEL* VNTR, which often results in low-quality sequence reads, especially for those repeats that are positioned

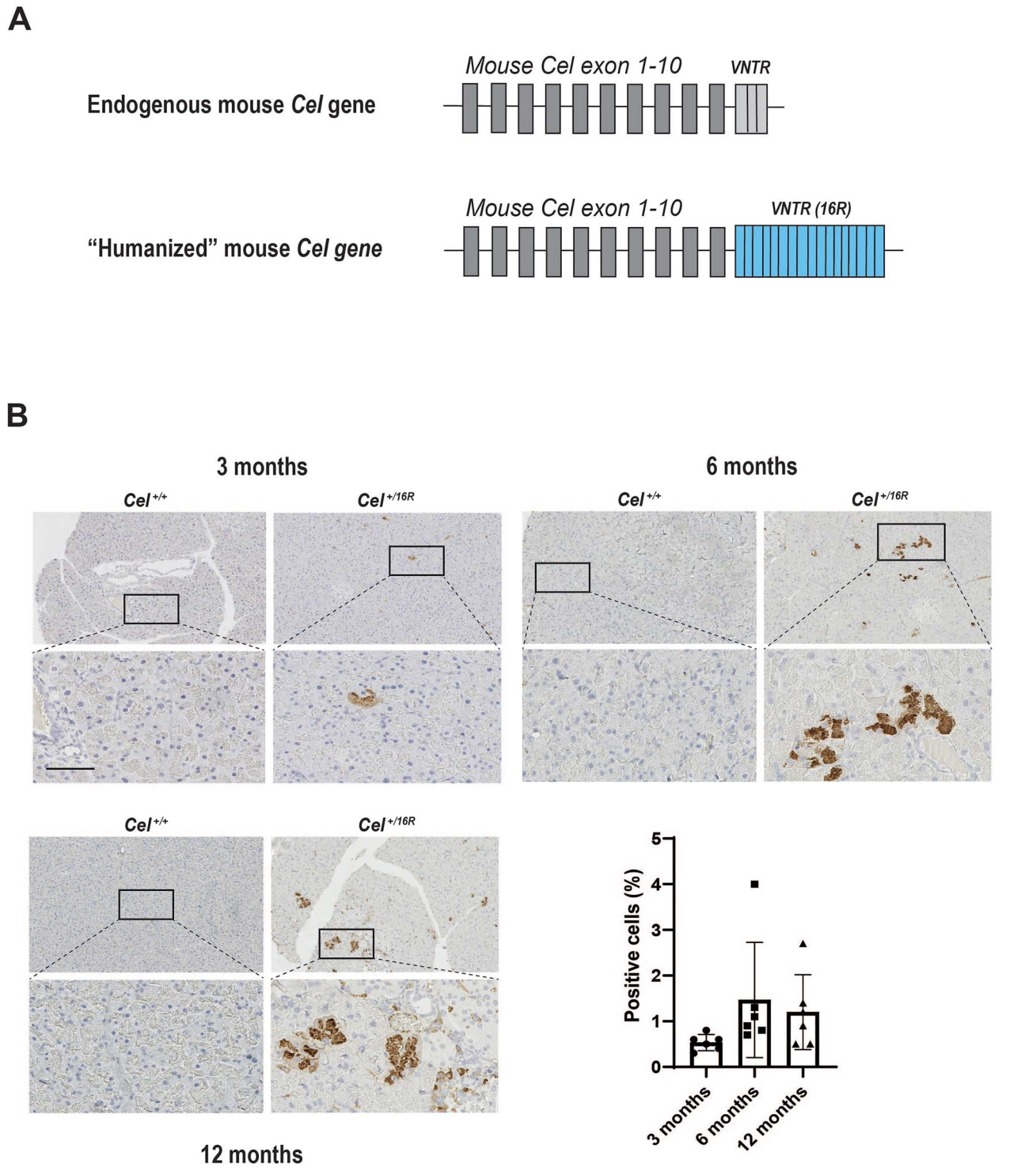

Figure 7. CEL-INS-positive staining in pancreatic tissue from a humanized knock-in mouse. (A) Schematic illustration of the *Cel* gene in the *Cel-16R* mouse strain, in which the VNTR of mouse *Cel* (three repeats) has been replaced by the VNTR of the normal human *CEL* gene (16 repeats = 16R). (B) Mouse pancreatic tissue sections stained with the anti-PRAAHG antibody. Representative staining (low and high magnification) is shown for ages 3, 6 and 12 months for wild type mice ($Cel^{+/+}$) and mice heterozygous for the human *CEL* VNTR ($Cel^{+/16R}$). There were n = 6 male mice in each group. Scale bar = 80 μm for the high-magnification images.

in the middle of the VNTR (Fig. 2A). In our hands, the *CEL* VNTR region needs to be amplified by long-PCR, followed by Sanger sequencing and manual inspection of the electropherograms, which is a resource-demanding approach for large materials. Herein, we analyzed a cohort of 200 Norwegian subjects with benign or malignant tumors of the pancreas. We found that 14.0% were carriers of various *CEL-INS* variants and that *INS9* was the by far most common allele. Tested carrier frequencies were very similar to the distribution in the control cohort. Thus, although our materials are of limited size, we found no evidence for an impact of germline *CEL-INS* variants on the risk for neoplastic pancreatic disease.

During our screening, we made the interesting observation that the recurrent *CEL-INS9* variant was always present on a VNTR allele with 13 repeats. Conversely, we noted that 75% of samples with a 13-repeat allele carried the *CEL-INS9* variant. These observations make it possible to check for effects of *CEL-INS9* by determining the distribution of samples with VNTR length 13, an assay that is technically simple [12]. In four studies that specifically have looked at pancreatic disease risk and *CEL* VNTR length, the 13-repeat allele was not identified as a risk factor [9, 10, 12, 13].

Still, single-base insertions can theoretically arise in any VNTR repeat, with different impact on the tail length of the translated CEL protein. The functional effect of an insertion may therefore vary according to its position within the VNTR. We have previously demonstrated such a pattern for single-base deletion mutations in the *CEL* VNTR: Deletions located in the most proximal repeats resulted in CEL protein variants that—compared to the normal protein—were less secreted from the cells and had an increased tendency to aggregate [24].

Hence, we set out to investigate if there is a similar positional effect among CEL-INS protein variants. Only CEL-INS1 exhibited significantly lower secretion than CEL-WT and were more abundant in the whole-cell lysate (Fig. 4A and B). In the soluble lysate fraction, the strongest reduction was noted for INS1, INS4 and INS7 (Fig. 4C). In the pellet fraction, all variants were significantly more abundant than CEL-WT, except INS12 (Fig. 4D). When assessing ER stress and UPR in the insoluble pellet, all markers were generally upregulated for INS4 and, in particular, for INS1.

Overall, the results depicted in Fig. 4 and Fig. 5 indicate that CEL-INS variants with insertions in the most proximal VNTR repeats have the highest propensity to form insoluble aggregates. When the point of insertion is shifted more towards the 3′ end of the VNTR, the CEL-INS protein behaves more similarly to CEL-WT. This indicates that a certain number of normal VNTR repeats is necessary for the CEL protein to be effectively secreted and to avoid aggregation. Moreover, our results are fully consistent with the findings of [24], where the effect of single-base deletions in the *CEL* VNTR became less severe as the point of mutation was positioned more distally in the VNTR.

Given their modest functional effects and relatively high population frequencies, we conclude that *CEL-INS9*, *CEL-INS10* and *CEL-INS12* are benign genetic variants. *CEL-INS7*, which we have yet to observe in any material, also appears to be a relatively mild variant. In contrast, the *CEL-INS1* variant (not observed so far) definitely could be pathogenic based on the experimental data presented here. Regarding *CEL-INS4*, it is the variant with the second strongest effects, and there was one carrier in our control material. We have also observed *INS4* twice in clinical materials: in a person with exocrine dysfunction [17] and in a person with diabetes (K. Fjeld, unpublished data). Notably, this single-base insertion within the 4th VNTR repeat will lead to a very short tail of the CEL protein with only three normal repeats. This is very reminiscent of the rare, truncated *CEL* variant observed in a Danish diabetes family [9]. In addition, *CEL-HYB1*, which is a genetic risk factor for developing chronic pancreatitis [14], only harbors three VNTR repeats. Therefore, we conclude that the *CEL-INS4* allele may have pathogenic potential and at present should be regarded as a variant of uncertain significance (VUS).

Contrasting the cellular data, lipase activity was largely unaffected in our set of CEL-INS variants when compared with normal CEL (Fig. 3). We find it unlikely that any modest difference in CEL activity will be of biological significance. In fact, a study of full-body *Cel* knock-out mice revealed no spontaneous exocrine or endocrine pancreatic phenotype, even if there was a complete loss of the enzyme [31]. As summarized in the Discussion of [24], there is now compelling evidence that pathogenic CEL variants do not initiate disease due to loss of function, but rather through protein aggregation and cytotoxicity.

Finally, we employed CEL tail-specific antibodies to assess expression of CEL-INS protein directly in pancreatic tissue. Surprisingly, we discovered that a pancreatic tissue section from an individual without any insertion variant in the germline still exhibited small areas of reactivity towards the anti-PRAAHG antibody (Fig. 6A). Martinez *et al.* [25] reported similar findings for pancreatic cancer patients without germline *CEL-INS* variants, and proposed that somatic, single-base insertions in the VNTR region of *CEL* could be associated with early pancreatic cancer development and even serve as a diagnostic tool. We therefore decided to test the antibody on pancreatic tissue from four *CEL-INS*-negative patients without malignant pancreatic disease. In all four cases, small anti-PRAAHG-positive areas or single cells were observed in morphologically normal pancreatic sections (Fig. 6B). One explanation is that the INS-positive areas could represent cells in which transcription of *CEL* mRNA is particularly prone to errors. An alternative and perhaps more likely hypothesis is that single-base insertions frequently arise spontaneously in the *CEL* VNTR as somatic DNA mutations. In either case, the events are not restricted to malignant conditions.

The potential somatic formation of CEL-INS-positive areas was confirmed in *Cel-16R* animals, a newly constructed mouse strain where the endogenous mouse *Cel* VNTR with three repeats has been substituted by the 16-repeat VNTR sequence dominating in human *CEL*. In pancreatic sections from the knock-in mice, but not from control animals, we noted small patches or single cells displaying anti-PRAAHG-positive staining (Fig. 7B). We therefore conclude that pancreatic tissue undergoes some type of somatic event through which single-base insertions in the *CEL* VNTR arise. The VNTR region could be prone to mutations because of its GC-rich, repetitive nature with several homopolymeric stretches. The *CEL* VNTR may therefore be considered hypermutable, a property that is likely to underlie the high degree of germline polymorphism characteristic for the gene.

Nevertheless, a limitation of our study is that we have not directly identified the postulated, somatic *CEL-INS* mutations. Although theoretically possible to detect via deep sequencing of DNA or RNA isolated from pancreatic tissue, we consider such an effort an arduous task. The normal human *CEL* VNTR consists of 16 almost identical 33-bp repeats, each repeat having a GC-content of 85% and homopolymeric tracts. It is challenging to sequence this region with confidence (see Fig. 2A). It is expected to be even more difficult to map short sequence reads back to the right repeat position within the VNTR, especially for the potentially mutated sequences found in only around 1% of the tissue area (Fig. 7B).

Since we are not able to determine the postulated mutations directly, they could be located more proximally in the VNTR than the germline variants and therefore have some pathogenic effect. There were, however, no apparent histological differences between the patchy INS-positive areas and the surrounding INS-negative pancreatic parenchyma (Fig. 6B).

In summary, we find that single-base insertions of the *CEL* VNTR are carried by around 15% of the Norwegian population. Interestingly, we also postulate that insertion variants are likely to emerge in the acinar parenchyma by somatic events. Given the presence of clusters of INS-positive cells in all samples examined, these events are likely to be frequent. Our observations, along with

data from the cellular studies, lead us to conclude that the most distal and common insertion variants (*CEL-INS9*, *INS10* and *INS12*) are benign. Nevertheless, the *CEL-INS4* variant and any insertions positioned more proximally in the *CEL* VNTR have pathogenic potential. Further studies are required to determine whether the proximal insertion variants are associated with disease.

Materials and methods

Patient samples

DNA samples from patients with pancreatic neoplastic disease (n = 95 males, n = 105 females) were from a biobank at Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen, Norway. The biobank consists of biological samples from patients referred to the Department of Gastrointestinal Surgery, Haukeland University Hospital, Bergen, Norway [28, 29]. The study was approved by the Research Ethics Committee of Western Norway (2013/1772), and the patients had given their informed consent. *CEL* VNTR lengths of the samples in the present study were determined in a previous report [12]. The control material consisted of DNA samples from 416 non-diabetic subjects, randomly selected among the controls used by Hertel *et al.* [32]. These controls had been drawn from an extensive, population-based study in mid-Norway (HUNT2), considered to be representative for the country as a whole [33].

Screening for insertion variants in the *CEL* VNTR

Screening was carried out by a long-range PCR with genomic DNA isolated from blood leukocytes as template. A product covering *CEL* exons 8–11 was amplified and the VNTR in exon 11 Sanger-sequenced. PCR was performed in a reaction volume of 14 μl containing 6.3 μl 2xGC buffer I (TaKaRa), 2 μl dNTP mix (TaKaRa), 0.38 μl of each primer L11F (20 μM) and VNTR-R (20 μM), 2.5 μl 5 M betaine solution (Sigma-Aldrich), 0.06 μl 5 U/μl LaTaq (TaKaRa), 1.38 μl ddH_2O and 1 μl 10 ng/μl genomic DNA. Amplification started with denaturation at 94°C for 1 min followed by 14 cycles of 94°C for 20 s and 60°C for 10 min; then 20 cycles of 94°C for 20 s and 62°C for 10 min, and a final elongation step of 72°C for 10 min, followed by cooling to 4°C.

Five μl of the PCR product was treated with 2 μl of Illustra ExoProStar (Sigma-Aldrich) by incubating for 15 min at 37°C followed by 15 min at 80°C. Sequencing was then performed in a total volume of 10 μl containing 2 μl treated template mix, 1 μl primer EF (5 μM) or VNTR-R (5 μM), 2 μl 5 M betaine solution, 2 μl 5X Sequencing buffer (Thermo Fisher Scientific), 1 μl Big Dye v. 3.1 (Thermo Fisher Scientific) and 2 μl ddH_2O. Amplification started with denaturation at 96°C for 10 min followed by 25 cycles of 96°C for 10 min and 58°C for 5 s, and a final elongation step of 60°C for 4 min, followed by cooling to 4°C.

Primer sequences were: L11F, 5'-GTGCCTCACTCATTCTTCTATG GCAAC-3'; VNTR-R, 5'-TCCTGCAGCTTAGCCTTGGG-3'; EF, 5'-CACACACTGGGAACCCT-3'.

Plasmid constructs and nomenclature

The expression vectors pcDNA3/CEL-16R and pcDNA3/V563X (CEL-TRUNC) were generated as previously described [23, 24]. CEL-TRUNC is an artificial truncation mutant where a stop codon has been introduced early in the first VNTR repeat, implying that CEL-TRUNC lacks the entire VNTR region except for the first four amino acids. By a mutagenesis service at GeneScript (Piscataway, NJ, USA), pcDNA3/CEL-16R was used as template to create plasmid constructs encoding the CEL insertion variants INS1, INS4, INS7, INS9, INS10, and INS12 (Fig. 1). This variant designation reflects the position of the single-base insertion within the 1st, 4th, 7th, 9th, 10th or 12th repeat of the VNTR. All plasmid constructs were verified by Sanger sequencing. Variant designation according to HGVS nomenclature is given in Table 2.

Cell culture

HEK293T (ATCC) and HEK293 (Clontech) cells were maintained at 37°C in a 5% CO_2 humidifier incubator and cultured in Dulbecco's Modified Eagle Medium with high glucose (4500 mg/l) supplemented with 10% fetal bovine serum (FBS). HEK293T cells were complemented with 100 U/ml penicillin and 100 U/ml streptomycin, whereas HEK293 cells were complemented with 100 U/ml Antibiotic Antimycotic (Thermo Fisher Scientific) in the medium.

Transfection of HEK293T cells

For purification of recombinant CEL proteins for enzyme activity assessment (Fig. 3), the cells were transfected with polyethyleneimine (23966-2, Polysciences) as previously described with a slight modification [24]. In brief, in each T75 cm^2 culture flask, cells were grown at a density of 5 × 10^6 cells in 12 ml growth medium with 5% FBS and transfected with 18 μg plasmid DNA mixed with 54 μg polyethyleneimine in 1.2 ml Opti-MEM media (Thermo Fischer Scientific). The growth medium was switched to 12 ml Opti-MEM 24 h after transfection. Medium was changed and harvested every 24 h for three consecutive days, then pooled and filtered through a 0.22 μm filter before protein purification.

For the CEL protein distribution and ER stress analysis (Fig. 4; Fig. 5), HEK293T cells were transfected using FuGene 6 (Promega) as previously described [23, 34]. Cells were seeded in 6-well plates at a density of 1 × 10^6 cells per plate in 2 ml growth medium containing 10% FBS. They were transfected with 1.67 μg plasmid DNA mixed with 5 μl FuGene 6 diluted in 100 μl Opti-MEM. The conditioned medium was removed 48 h post-transfection, and the cells were supplied with 1.2 ml Opti-MEM. Sample collection was carried out 20–24 h after the medium change.

Transfection of HEK293 cells

For testing of the CEL tail-specific antibodies anti-PAVIRF and anti-PRAAHG (Supplementary Fig. 2), HEK293 cells were transfected using Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's instructions. Cells were seeded in 6-well plates (1 × 10^5 cells/well) and grown for 72 h for western blotting, or in a 12-well plate (2.5 × 10^5 cells/well) and grown for 24 h for immunofluorescent staining. The cells were transfected with plasmids encoding CEL-WT, CEL-INS9, CEL-TRUNC, or empty pcDNA3.1 vector as negative control.

Protein purification

CEL protein variants were purified from the conditioned media by a 5-ml HiTrap Heparin HP Affinity Column controlled by an ÄKTA Pure 25 chromatography system (GE Healthcare) as described previously [35]. For each CEL variant, concentration of the purified protein was determined by measuring absorbance at 280 nm using the corresponding extinction coefficient. To assess the integrity of the recombinant CEL proteins, 5 μg of each purified CEL variant were loaded and resolved on 4%–15% Mini-Protean TGX gels (Bio-Rad). The gels were then stained with GelCode Blue Stain Reagent (Thermo Fisher Scientific) for visualization of protein bands. Gel images were acquired using ChemiDoc MP Imaging System (BioRad) and the densitometry analysis of protein bands was performed using Image lab software (Bio-Rad).

Table 2. Overview of the investigated *CEL-INS* variants.

Variant	Nucleotide change[a]	Nucleotide change[b]	Amino acid change[c]
INS1	c.1680dupC	c.1671dupC	p.Thr558Hisfs*19
INS4	c.1785dupC	c.1776dupC	p.Val593Argfs*6
INS7	c.1884dupC	c.1875dupC	p.Val626Argfs*6
INS9	c.1950dupC	c.1941dupC	p.Val648Argfs*6
INS10	c.1983dupC	c.1974dupC	p.Val659Argfs*6
INS12	c.2049dupC	c.2040dupC	p.Val681Argfs*6

[a]Position according to previous NCBI reference mRNA sequence NM_001807.4. [b]Position according to the most recent NCBI reference mRNA sequence NM_001807.6. [c]Positions according to NCBI protein reference sequence NP_001798.3. Variant designation is according to Human Genome Variation Society (HGVS) nomenclature Reference sequences can be found at www.ncbi.nlm.nih.gov.

Lipase activity assay

The lipase activity assay was carried out using a 5-minute pH-stat method with trioctanoin (T9126, Sigma-Aldrich) in the presence of 12 mM sodium cholate (C1254, Sigma-Aldrich) as described previously [24]. For each assay reaction, 500 μl conditioned media or 7.5 μg purified CEL variant, was added. Lipase activities were expressed as units per mg protein (1 unit corresponds to 1 μmol of fatty acid released/min). The activities were adjusted according to densitometry and to the molecular weight of each variant.

Cellular fractionations

HEK293T cells were harvested 68–74 h after transfection (20–24 h after medium change) as previously described [34]. Conditioned medium was collected and analyzed as the "medium" fraction. Ice-cold PBS was added to the transfected cells, and the cell suspension was collected and centrifuged at 500 × g for two min. The supernatant was removed, and the cell pellet was washed twice in PBS. Next, the cells were lysed in 500 μl radio-immunoprecipitation buffer (Sigma-Aldrich) supplemented with EDTA-free complete protease inhibitor mixture (Roche) and incubated on ice for 10 min, followed by a 3 × 10 s sonication. To prepare the "whole-cell lysate", 75 μl of the sonicated cell lysate was mixed with an equal volume of 2 × Laemmli buffer and boiled at 95°C for 10 min, following a 3 × 10 s sonication. The remaining of the initial cell lysates were centrifuged at 16 000 × g for 20 min at 4°C. The resulting supernatants will be referred to as the "soluble lysate" fraction. The remaining pellet was washed twice with ice-cold PBS before centrifugation at 16 000 × g for 5 min, resuspension in 100 μl 1 × Laemmli buffer and sonication for 3 × 10 s. This preparation was designated the "insoluble pellet" fraction.

Western blotting

Western blotting was carried out by standard methods and as previously described [24]. Primary antibodies were: anti-BiP, 1:1000 (3183S); anti-calnexin, 1:2000 (2679S); anti-PERK, 1:1000 (3192S); anti-calreticulin, 1:1000 (12238S)—all rabbit polyclonal antibodies from Cell Signaling Technology. Moreover, we employed a rabbit polyclonal antibody against the globular domain of CEL at 1:10000 [23]; rabbit polyclonal antibodies against the six last amino acids of CEL-WT (anti-PAVIRF) or CEL-INS (anti-PRAAHG) at 1:5000 [25]; mouse monoclonal anti-GAPDH (sc-47 724, Santa Cruz Biotechnology) at 1:1000, rat monoclonal anti-α-tubulin at 1:1000 (sc-53 029, Santa Cruz Biotechnology), and mouse monoclonal anti-β-actin at 1:1000 (A5441, Sigma-Aldrich).

Secondary antibodies were: IRDye 680 goat-anti-rabbit IgG (D20601-15, LI-COR; 1:10000), IRDye 800 goat anti-rat IgG (D01118-15, LI-COR; 1:10000); goat anti-rabbit IgG (65-6120, Invitrogen; 1:5000 or 1:10000), donkey anti-mouse IgG (sc-2318, Santa Cruz Biotechnology; 1:5000) and goat anti-mouse IgG (62-6520, Invitrogen; 1:5000). When needed, blots were developed using Pierce ECL Plus Western Blotting Substrate (Thermo Fisher Scientific), and signals were analyzed using a Las 1000 Pro v. 2.6 software imager (Fujifilm) or ChemiDoc MP Imaging System (BioRad). Band quantification was performed using the Image Gauge software v4.0 software (Fujifilm) or Image Lab (BioRad).

Immunofluorescent staining and confocal imaging

Immunocytochemistry was carried out as previously described [36]. Primary CEL antibodies and dilutions were As20.1 (1:200), and anti-PRAAHG (1:5000). As20.1 is a mouse monoclonal antibody against the CEL globular domain, kindly provided by Prof. Olle Hernell, Department of Clinical Sciences, Umeå University, Sweden. Secondary antibodies were donkey anti-rabbit Alexa Fluor 488 (A-21206, Thermo Fisher Scientific) or anti-mouse Alexa Fluor 488 (A-11017, Thermo Fisher Scientific), both at 1:200. The coverslips were mounted on glass slides using Prolong Gold Antifade Solution with 4′,6-diamidine-2′-phenylindole dihydrochloride (DAPI). Images were acquired with a TCS SP5 confocal microscope (Leica Microsystems) using Diode 405 and Argon 488 lasers. Series of images were obtained at 0.25-μm intervals throughout the entire depth of the cells.

Immunohistochemistry

Immunohistochemistry was carried out as previously described [37]. Prior to adding the primary antibody, the slides were treated with serum-free protein block (X0909, Agilent Dako). Primary antibodies and dilutions were anti-PAVIRF (1:5000), anti-PRAAHG (1:5000), and anti-CEL (HPA052701, Sigma-Aldrich; 1:100-200). Primary antibody detection was done by MACH3 anti-rabbit probe and HRP-conjugated polymers (Biocare Medical), followed by visualization with 3,3′-diaminobenzidine (Dako). The sections were counterstained with hematoxylin and images acquired by using a Leica DMLB microscope with the ZEN 2011 software. Alternatively, the sections were scanned using (Nano Zoomer XR, Hamamatsu Photonics) and images obtained by Aperio ImageScope (Leica Biosystems). The software QuPath (qupath.github.io) was used for quantification of INS-positive cells in mouse pancreatic sections.

Generation of the *Cel-16R* mouse

The *Cel-16R* knock-in mouse strain was established on a C57Bl/6N genetic background by GenOway (Lyon, France). The strain was constructed similarly to the *Cel-HYB1* mouse described in [38], except that the VNTR in mouse *Cel* exon 11 was substituted with the VNTR of the normal human *CEL* gene (consisting of 16 repeats). Verified animals were transferred to the University of Bergen, Norway and used to establish a colony of *Cel-16R* mice

by backcrossing with C57Bl/6N wild-type mice from Charles River Laboratories (Lyon, France). The animals were housed on a 12-h light/dark cycle with ad libitum access to water and food at the Laboratory Animal Facility, Faculty of Medicine, University of Bergen. Sampling of pancreatic tissue was approved by the Norwegian Animal Research Authority (FOTS 13510). Mice were euthanized by CO_2 and the pancreas immediately collected and fixated in 10% buffered formalin for histology. Genotyping of *Cel-16R* mice was performed using forward primer 5′-GCA AAC TTC TTA TTT ATC CTC AAG CC TTG G-3′ and reverse primer, 5′-GTT ATC GTC TTA GTG ATG TCC AGG TAG TTG C-3′. Amplicon sizes from the wild-type *Cel* and *Cel-16R* alleles were 303 and 394 bp, respectively (Supplementary Fig. 4A).

Statistical analysis

Results are given as the mean ± standard deviation (SD). Ordinary one-way ANOVA was used for statistical analysis, except for assessing carrier frequencies of *CEL-INS* variants, which was done by a 2×2 contingency table and two-tailed Chi-square test. Testing was carried out using GraphPad Prism (www.graphpad.com) or the webpage www.quantitativeskills.com/sisa. *P*-values ≤ 0.05 were considered statistically significant.

Supplementary data

Supplementary data is available at *HMG Journal* online.

Conflict of interest statement: The authors declare no conflict of interests.

Funding

R.S.B. was supported by a PhD fellowship from the University of Bergen. Financial support for this study came from Research Council of Norway (FRIMEDBIO #289534), Western Norway Regional Health Authority (Helse Vest #912057) and Norwegian Cancer Society (Kreftforeningen #212734-2019; #245030-2022) to A.M., from National Institutes of Health (R01 DK124415/NIDDK) to M.E.L. and from National Pancreas Foundation to X.X.

References

1. Whitcomb DC, Lowe ME. Human pancreatic digestive enzymes. *Dig Dis Sci* 2007;**52**:1–17.
2. Blackberg L, Lombardo D, Hernell O. *et al.* Bile salt-stimulated lipase in human milk and carboxyl ester hydrolase in pancreatic juice: are they identical enzymes? *FEBS Lett* 1981;**136**:284–8.
3. Lombardo D, Guy O. Studies on the substrate specificity of a carboxyl ester hydrolase from human pancreatic juice. II. Action on cholesterol esters and lipid-soluble vitamin esters. *Biochim Biophys Acta* 1980;**611**:147–55.
4. Hui DY, Howles PN. Carboxyl ester lipase: structure-function relationship and physiological role in lipoprotein metabolism and atherosclerosis. *J Lipid Res* 2002;**43**:2017–30.
5. Lombardo D. Bile salt-dependent lipase: its pathophysiological implications. *Biochim Biophys Acta* 2001;**1533**:1–28.
6. Johansson BB, Fjeld K, El Jellas K. *et al.* The role of the carboxyl ester lipase (CEL) gene in pancreatic disease. *Pancreatology* 2018;**18**:12–9.
7. Higuchi S, Nakamura Y, Saito S. Characterization of a VNTR polymorphism in the coding region of the *CEL* gene. *J Hum Genet* 2002;**47**:213–5.
8. Lindquist S, Blackberg L, Hernell O. Human bile salt-stimulated lipase has a high frequency of size variation due to a hypervariable region in exon 11. *Eur J Biochem* 2002;**269**:759–67.
9. Torsvik J, Johansson S, Johansen A. *et al.* Mutations in the VNTR of the carboxyl-ester lipase gene (*CEL*) are a rare cause of monogenic diabetes. *Hum Genet* 2010;**127**:55–64.
10. Fjeld K, Beer S, Johnstone M. *et al.* 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* 2016;**11**:e0165567.
11. Fjeld K, Masson E, Lin JH. *et al.* Characterization of *CEL-DUP2*: complete duplication of the carboxyl ester lipase gene is unlikely to influence risk of chronic pancreatitis. *Pancreatology* 2020;**20**:377–84.
12. Dalva M, El Jellas K, Steine SJ. *et al.* Copy number variants and VNTR length polymorphisms of the carboxyl-ester lipase (*CEL*) gene as risk factors in pancreatic cancer. *Pancreatology* 2016;**17**:83–8.
13. Mao XT, Deng SJ, Kang RL. *et al.* Homozygosity of short VNTR lengths in the *CEL* gene may confer susceptibility to idiopathic chronic pancreatitis. *Pancreatology* 2021;**21**:1311–6.
14. Fjeld K, Weiss FU, Lasher D. *et al.* A recombined allele of the lipase gene *CEL* and its pseudogene *CELP* confers susceptibility to chronic pancreatitis. *Nat Genet* 2015;**47**:518–22.
15. Oracz G, Kujko AA, Fjeld K. *et al.* The hybrid allele 1 of carboxyl-ester lipase (*CEL-HYB1*) in Polish pediatric patients with chronic pancreatitis. *Pancreatology* 2019;**19**:531–4.
16. Tjora E, Gravdal A, Engjom T. *et al.* Protein misfolding in combination with other risk factors in *CEL-HYB1*-mediated chronic pancreatitis. *Eur J Gastroenterol Hepatol* 2021;**33**:839–43.
17. Ræder H, Johansson S, Holm PI. *et al.* Mutations in the *CEL* VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction. *Nat Genet* 2006;**38**:54–62.
18. Pellegrini S, Pipitone GB, Cospito A. *et al.* Generation of β cells from iPSC of a MODY8 patient with a novel mutation in the carboxyl ester lipase (CEL) gene. *J Clin Endocrinol Metab* 2021;**106**:e2322–33.
19. El Jellas K, Dusatkova P, Haldorsen IS. *et al.* Two new mutations in the *CEL* gene causing diabetes and hereditary pancreatitis: how to correctly identify MODY8 cases. *J Clin Endocrinol Metab* 2022;**107**:e1455–66.
20. Ræder H, McAllister FE, Tjora E. *et al.* 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* 2014;**63**:259–69.
21. Johansson BB, Torsvik J, Bjørkhaug L. *et al.* Diabetes and pancreatic exocrine dysfunction due to mutations in the carboxyl ester lipase gene-maturity onset diabetes of the young (CEL-MODY): a protein misfolding disease. *J Biol Chem* 2011;**286**:34593–605.
22. Torsvik J, Johansson BB, Dalva M. *et al.* Endocytosis of secreted carboxyl ester lipase in a syndrome of diabetes and pancreatic exocrine dysfunction. *J Biol Chem* 2014;**289**:29097–111.
23. Xiao X, Jones G, Sevilla WA. *et al.* A carboxyl ester lipase (CEL) mutant causes chronic pancreatitis by forming intracellular aggregates that activate apoptosis. *J Biol Chem* 2017;**292**:7744–4.
24. Gravdal A, Xiao X, Cnop M. *et al.* The position of single-base deletions in the VNTR sequence of the carboxyl ester lipase (*CEL*) gene determines proteotoxicity. *J Biol Chem* 2021;**296**:100661.
25. Martinez E, Crenon I, Silvy F. *et al.* Expression of truncated bile salt-dependent lipase variant in pancreatic pre-neoplastic lesions. *Oncotarget* 2017;**8**:536–51.

26. Genomics England, London. *Monogenic Diabetes Panel (Version 2.54)*. Available from: https://panelapp.genomicsengland.co.uk/panels/472/ (Downloaded 13.02.2024).
27. Genomics England, London. *Pancreatitis Panel (Version 3.4)*. Available from: https://panelapp.genomicsengland.co.uk/panels/386/ (Downloaded 13.02.2024).
28. El Jellas K, Hoem D, Hagen KG. *et al.* Associations between ABO blood groups and pancreatic ductal adenocarcinoma: influence on resection status and survival. *Cancer Med* 2017;**6**:1531–40.
29. Choi MH, Mejlænder-Andersen E, Manueldas S. *et al.* Mutation analysis by deep sequencing of pancreatic juice from patients with pancreatic ductal adenocarcinoma. *BMC Cancer* 2019;**19**:11.
30. Karczewski KJ, Weisburd B, Thomas B. *et al.* The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res* 2017;**45**:D840–5.
31. Vesterhus M, Ræder H, Kurpad AJ. *et al.* Pancreatic function in carboxyl-ester lipase knockout mice. *Pancreatology* 2010;**10**:467–76.
32. Hertel JK, Johansson S, Sonestedt E. *et al. FTO*, type 2 diabetes, and weight gain throughout adult life: a meta-analysis of 41,504 subjects from the Scandinavian HUNT, MDC, and MPP studies. *Diabetes* 2011;**60**:1637–44.
33. Holmen J, Midthjell K, Krüger Ø. *et al.* The Nord-Trøndelag health study 1995-97 (HUNT 2): objectives, contents, methods and participation. *Norsk Epidemiologi* 2003;**13**:19–32.
34. Cassidy BM, Zino S, Fjeld K. *et al.* Single nucleotide polymorphisms in *CEL-HYB1* increase risk for chronic pancreatitis through proteotoxic misfolding. *Hum Mutat* 2020;**41**:1967–78.
35. Xiao X, Mukherjee A, Ross LE. *et al.* Pancreatic lipase-related protein-2 (PLRP2) can contribute to dietary fat digestion in human newborns. *J Biol Chem* 2011;**286**:26353–63.
36. Dalva M, Lavik IK, El Jellas K. *et al.* Pathogenic carboxyl ester lipase (CEL) variants interact with the normal CEL protein in pancreatic cells. *Cells* 2020;**9**:244.
37. El Jellas K, Johansson BB, Fjeld K. *et al.* The mucinous domain of pancreatic carboxyl-ester lipase (CEL) contains core 1/core 2 O-glycans that can be modified by ABO blood group determinants. *J Biol Chem* 2018;**293**:19476–91.
38. Fjeld K, Gravdal A, Brekke RS. *et al.* The genetic risk factor *CEL-HYB1* causes proteotoxicity and chronic pancreatitis in mice. *Pancreatology* 2022;**22**:1099–111.