SMART Topic

Six Types of Pancreatitis Risk Factors in the Trypsinogen Gene Loci (*PRSS1-PRSS2-TRB*).

Andrea Geisz PhD[1], Maham Waqar MD[2,3] and David C. Whitcomb MD PhD[2.4]

[1]Department of Surgery, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, [2]Ariel Precision Medicine, Pittsburgh, PA, [3]Allegheny General Hospital Internal Medicine Department, Pittsburgh, PA, . [4]Department of Medicine, University of Pittsburgh, Pittsburgh, PA.

Abstract:

Trypsin dysregulation is central to acute pancreatitis (AP) and contributes to chronic pancreatitis (CP). Genetic variants within the T Cell receptor beta chain (*TRB*) locus, that includes the trypsinogen genes (*PRSS1* and *PRSS2*) affect risks of AP and CP. Hereditary pancreatitis is caused by rare *gain-of-function* variants (e.g. *PRSS1* p.N29I, p.R122H), while other common and rare variants alter risk for AP and/or CP through 5 other mechanisms, including altering the *TRB* gene receptor repertoire. Here we provide a brief primer of genetic risk for AP and CP using examples of variants within the *TRB* locus and the *PRSS1* and *PRSS2* genes.

I. Background

Multiple lines of evidence link acute pancreatitis (AP), recurrent AP (RAP), and chronic pancreatitis (CP) to the pancreatic digestive enzyme, **trypsin** (below). Trypsin is the master digestive enzyme that functions as the catalyst for activating all other pancreatic pro-enzymes (zymogens). Trypsin is normally activated from trypsinogen by a duodenal mucosal enzyme, enterokinase (Transmembrane Serine Protease 15, *TMPRSS15*) in the small intestine or activated by another active trypsin molecule, a reaction commonly called *auto-activation.*

Trypsin is responsible for acute pancreatitis.

The discovery that the gain-of-function *PRSS1* p.R122H mutation caused hereditary pancreatitis (HP) indicated that trypsin activation in the pancreas was an important factor in AP. [1, 2] The centrality of trypsin in AP is highlighted by genetically engineered mouse models (GEMMs). GEMMs with trypsinogen gene knockout do not develop AP, but they *can* develop CP with chronic acinar cell hyperstimulation [3]. On the other hand, GEMMs with knockout of mouse trypsin gene and insertion of normal or mutated human trypsinogen gene (*PRSS1*) develop AP spontaneously or with minimal stress or injury.[4]

Similarly, knock-in GEMMS carrying rapidly autoactivating trypsinogen mutants exhibit spontaneous pancreatitis or increased sensitivity to experimentally induced pancreatitis.[5-8] [5-8]Furthermore, mice with more easily activating trypsinogens progress to CP as in human HP. [5, 7, 8]

The importance of trypsin in pancreatitis cannot be underestimated. In addition to disease-associated mutations in the trypsinogen genes, other genetic factors that protect the pancreas from trypsin are also important, but not discussed here. These include variants affecting trypsin inhibitors (e.g. *SPINK1, CTRC*), trypsin flushing (*CFTR*), and other factors.

> **Trypsin activation is the key to acute pancreatitis (AP). Excess, persistent or inhibition-resistant trypsin increases the risk of AP and then CP. CP can be driven independently of trypsin activation.**

II. Short Primer on Genetics

This primer is a short review of the effect of genetic variants on pancreatic diseases, focusing on trypsin biology. Readers familiar with genetics may choose to skip to Section III for a discussion of the 6 types of trypsin and *TRB* variants that are relevant to the biology of pancreatitis. See also the Human Genomic Variation Fact Sheet, National Human Genome Research Institute.

DNA is made up of 4 nucleotide bases (ATGC). A kilobase (kb) refers to a DNA sequence of 1000 base pairs. Genetic variants can be found in **exons** (the protein coding DNA with three base **codons** coding for amino acids (AA)), the regions between exons (**introns**) and upstream or downstream of the coding exons, in the immediate 5' and 3' regions or at longer distances from a gene (**intergenic** regions). Most genetic variants involve the substitution of one nucleotide for another called single nucleotide *variants* [SNVs] or *polymorphisms*

Abbreviations used in this paper. AA amino acid, AP, acute pancreatitis, CF cystic fibrosis, CNV copy number variants, CP, chronic pancreatitis; ER endoplasmic reticulum, GEMMs genetically engineered mice models, MAF, minor allele frequency; PRS, polygenic risk score, RAP. recurrent acute pancreatitis, SNV single nucleotide variant, SNP single nucleotide polymorphism

ISSN 2997-2876 (online)
ISSN 2997-2868 (print)
DIO: https://doi.org/10.69734/x2yz2m74
Website: SMART-MD.org

[SNPs]. These are also called *point mutations*. Less common, but important are DNA sequence variants that span large regions such as deletions or insertions variants (**DELINs**), duplications (copy number variants, [**CNV**]), or more complex structural rearrangements (e.g. translocations, inversions, loss) or chromosome-level changes (e.g., Down's syndrome).

Medical Genetics typically focuses on genetic variants in the exons, or higher-level chromosomal deletions, duplication or rearrangement (e.g. Down's syndrome, Klinefelter syndrome). For single gene syndromes (e.g. hereditary pancreatitis, HP; cystic fibrosis, CF) the effect of a mutation in the coding region can range from minimal to damaging to normal function. Recessive traits (e.g. CF) usually involve *loss-of-function* variants affecting both alleles (chromosomes). Dominant traits (e.g. HP) typically involve *gain-of-function* variants on one allele and are rare. Additionally, some dominant variants are *loss-of-function* variants in genes where expression of both alleles is needed for normal function.

Mendelian Genetics focuses on familial inheritance patterns and monogenetic diseases (one gene-one disease). Less than 5% of human pancreatitis is due to monogenic causes. The rest are complex polygenic risks that become pathogenic with weak or strong environmental or metabolic risk factors.

*Variants Altering Protein **Function**.*

DNA sequence variants that alter protein function are typically located within the coding region of a gene (e.g. codon). Brief definitions and examples of some variants relevant to pancreatic disease are listed in **Table 1**.

Table 1. Seven types of genetic variants affecting the protein coding regions (codons) with pancreatitis examples.

- **Nonsense**: SNP changing an AA codon to a STOP codon– presumably protective in the context of *PRSS1* (e.g. rare variants p.Tyr37Ter, p.Gln56Ter) but typically pathogenic in other genes.
- **Frameshift**: Insertion or deletion of a nucleotide in a codon changing the number to >3, <3 changing the AA sequence (e.g. *CFTR* p.Gly330fs) - typically pathogenic.
- **Missense**: Substitution of a nucleotide within a codon to cause the wrong AA to be translated. This typically causes misfolding of the protein product that can trigger an unfolded protein response in the rough endoplasmic reticulum (e.g. *PRSS1* p.R116C) or alter the function of a fully formed protein (e.g. *PRSS1* p.R122H).
- **Synonymous**: SNPs that change the codon sequence but do not change the AA (since many amino acids are linked to more than one codon). These are not necessarily silent mutations as they may alter transcription factors (TF), alter folding, and expression or be linked to a pathogenic or protective haplotype (e.g. *PRSS1* p.N246N also denoted as p.N246=).
- **Splice site variants:** DNA sequence variants near the junction of introns and exons may cause an exon to be skipped, or intronic nucleotides to be included in mRNA and be translated. The *CFTR* "5T" allele may cause exon skipping and synthesis of dysfunctional CFTR. Examples in *PRSS1* are the rare variants c.40+1G>A and c.200+1G>A, which presumably protect against pancreatitis.
- **Deletion- Insertion** (DELIN) variants: These variants represent small insertions or deletions of DNA, typically <50 base pairs but can be much larger. Results can be nonsense or frameshift mutations, in-frame deletions (e.g. CFTR p.F508del), in-frame insertions (*PRSS1* p.K23_I24insIDK) or may include whole exons or genes. Relevant to pancreatitis is the 20kb DELIN initially dubbed the *PRSS1-PRSS2* haplotype (described below).
- **Copy Number Variants** (CNV). These are a type of sequence changes where, compared to the reference sequence, a copy of the entire gene (or other defined DNA sequence) is typically inserted directly 3' of the reference gene. A CNV is a type of **duplication** variant.

Medical geneticists classify variants in the coding region as *Benign*, *Likely Benign*, *unknown* (variant of unknown significance, **VUS**), *Likely Pathogenic* or *Pathogenic* based on correlation with computer models, biological experiments and clinical impact (levels of evidence).[9]

*Variants Altering Protein **Expression**.*

A small minority of genetic variants affect the protein coding region of genes. Most pancreatitis risk variants are in non-coding regions that affect gene expression either directly or indirectly (e.g. through coding or non-coding genes that regulate other genes). One example is **transcription factor** (TF) binding sites that regulate expression. Those sites that are close to the exons are typically on the same chromosome strand as the exons and are said to be in "*cis*", and typically affect only one chromosome. Those that are far away are said to be in "trans" and can affect the expression of the gene on both chromosomes. There are many examples of genetic variants in an exon that cause a synonymous amino acid change (the normal and variant codons code for the same amino acid) but the alternated sequence is pathogenic because it alters a key TF binding site.

Haplotypes. **Haplotypes** are very important in genetics. A haplotype refers to a set of DNA variants along a single chromosome that tend to be inherited together (see NCI example). Since there are usually many SNPs on a haplotype, it is often difficult to determine exactly which non-coding variant is causing a damaging effect on gene expression. Furthermore, in many cases multiple SNPs may have synergistic or additive effects.

Specialized Cells. Almost every cell in the body contains all the genetic information for all the genes in the organism. Thus, the issue of genetic variants that affect gene expression

exceeds transcription factors; it includes higher level DNA packaging, with some (most) genes being blocked from translation by histones and other factors. Secondly, there may be different specialized cells that express the same gene at different times and under different circumstances. One regulatory mechanism is linked to differences in the expression of TFs that initiate or block gene transcription. Thus, in some cases a gene is expressed normally in one tissue, but not in a second because the TF binding site that is key to gene expression has a sequence variant. Thus, the gene product is not expressed sufficiently under critical conditions to meet the cell's needs.

Most of the genetic variants identified in genome-wide association studies (GWAS) are in non-coding regions, or in coding regions with synonymous variants. Furthermore, these variants tend to be common, so they cannot *cause* the disease – but clearly contribute to it in some way. This is where various "omic" fields, such as *transcriptomics* are needed to define exactly what effect the variant (or haplotype) has on expression of one or more genes in one or more tissues under one or more conditions.

The key to precision medicine is understanding the effects of genetic variants on *protein function* (gain, loss or change) and *protein expression* (right amount, right time) in specific specialized cells or systems that can cause pathogenic responses to activation or stress, thereby leading to disease.

The importance of a good precision medicine report is that it considers not only the DNA sequence changes, but also what the integrated effects of all relevant variants in all relevant cells and systems mean in a patient with signs and symptoms, or risks of complications in a complex condition or disease.

III. Three trypsin(ogen) genes.

Trypsin is the master digestive enzyme of the pancreas. Three forms of trypsin are expressed: cationic (PRSS1), anionic (PRSS2), and mesotrypsin (PRSS3), which are the activation products of three pro-enzymes (zymogens) called trypsinogens. The trypsinogen genes *PRSS1* and *PRSS2* are located on chromosome 7 (*TRB* locus), whereas *PRSS3* is on chromosome 9. *PRSS1* has been reported to be expressed at about twice the level of *PRSS2*, and *PRSS3* generally represents less than 5% of all trypsinogens.

In addition to the 3 trypsin genes that are expressed in humans, there are a variable number of trypsinogen pseudogenes that have DNA sequences with high similarity to the expressed trypsinogens. This is important when evaluating DNA sequencing results so that the exact sequence of *PRSS1*, *PRSS2* and *PRSS3* are determined, especially in some key functional and regulatory regions.

Three trypsinogens *PRSS1, PRSS2*, and *PRSS3*, are encoded on chromosomes 7 and 9, with *PRSS1* predominantly expressed. Pseudogenes, like the expressed trypsinogens require careful differentiation during genetic analysis for accurate genotyping of the expressed trypsinogen gene sequences in each subject.

Table 2. Six *TRB* and trypsinogen gene locus variants affecting pancreatitis.

1) ***PRSS1* exon missense variants** with *gain-of-function* in the trypsin protein: (e.g. *PRSS1* p.N29I, p.R122H). These cause autosomal dominant HP with AP, RAP and CP (see warning about *PRSS1* p.N28I sequencing below).

2) **Copy number variants (CNV)** in the *PRSS1* and *PRSS2* genes. (e.g. 3 or more copies rather than 2) resulting in excess trypsinogen in the pancreas: This rare CNV results in autosomal dominant HP with AP, RAP and CP.

3) ***PRSS1* exon missense variants causing protein misfolding.** Some mutations cause *protein misfolding* and retention of the protein products in the RER "non-secreted" triggering a sustained unfolded protein response (UPR). CP can arise from the sustained stress response and associated inflammation. Autosomal dominant CP may develop with low penetrance. These mutations appear to be rare.

4) ***PRSS1 / PRSS2 loss-of-function AP protection variants*** with *loss-of-function* may be *protective* for the development of AP or CP. These variants, (e.g. *PRSS2* p.G191R) are more common in people without pancreatitis. The *PRSS2* missense variant p.Thr8Ile (rs62473563) may be another important example based on reduced *PRSS2* expression.[10]

5) ***PRSS1-PRSS2-TRB* Risk Haplotype.** A large, ~100kb haplotype that spans all of the trypsin genes and pseudogenes on chromosome 7 is associated with CP and is tagged by multiple SNPs including *PRSS1* p.N246N (rs6667_C). The genetic effect is complex and indirect, including increased expression of *PRSS2* (increased ~25%, see GTEx with search term "pancreas" under eQTL). It also has major effects on the T-cell receptor beta chain variable region (TRB) with large effects on T-cell activation and regulation[11-13] – with an observed effect of accelerating transition for RAP to CP, especially in the presence of alcohol abuse.[14]

6) ***PRSS1-PRSS2-TRB* Protective Haplotype.** This primary alternate 100kb haplotype in the PRSS1-PRSS2-TRB region is marked by a 20kb deletion, including two of the trypsinogen pseudogenes. This protective haplotype is tagged by *PRSS1* p.N246 (rs6667_T) and other SNPs.[10] As with the risk haplotype, the mechanisms that protect the pancreas from injury and inflammation are complex, but may include partial linkage to the *PRSS2* p.Thr8Ile (rs62473563) variant noted above (mechanism 4).

IV. Pancreatitis & the trypsinogen gene loci.

The clinical problem of pancreatitis is typically centered on altered regulation or activation of the trypsinogen genes. As noted in *Section II. Short Primer on Genetics*, there are multiple types of genetic variants that can affect gene function or expression. Each of these is important in human pancreatic diseases with different types of mutations having different effects and (likely) different treatment approaches.

1. HP from gain-of-function mutations

Hereditary pancreatitis is caused by "**gain-of-function**" mutations in the *PRSS1* gene coding regions including *PRSS1* p.N29I and p.R122H. In this class, gain-of-function is caused by impairment or delay in degradation of trypsinogen or active trypsin. Since only one copy of "poorly controlled" trypsin can drive the zymogen activation cascade, it is sufficient to cause autosomal dominant HP. The phenotype is recurrent episodes of acute pancreatitis (RAP). In most cases the recurrent damage will lead to features of CP including atrophy, fibrosis, diabetes and risk of pancreatic cancer.

> **HP is diagnosed from the family tree and confirmed with either *PRSS1* p.N29I or N29T, p.R122C or R122H. Systematic sequencing problems cause misdiagnosis of HP in patients with erroneous p.A16V and p.N29I genotype calls.**

2. PRSS1-PRSS2 copy number variant.

HP can also arise from duplicate/triplicate copy number variants (CNV) of *PRSS1 / PRSS2* genes. Multiple copies of *PRSS1* (and possibly *PRSS2*) result in overproduction of trypsinogen in the pancreas and increases risk of pancreatitis.

The CNV denotes a specific DNA segment of at least 1kb in size that is inserted multiple times compared to the reference genome. Heterozygous duplication and triplication of a 605-kb sequence containing both *PRSS1* and *PRSS2* on chromosome 7 has been reported in patients with hereditary and idiopathic chronic pancreatitis.[15-18] Pathogenicity was presumed to be the result of a gene dose effect due to the increased number of trypsinogen copies. This concept was later confirmed by mouse models, in which transgenically expressed wild type human trypsinogen impacted disease phenotype in a dose-dependent manner.[19-21] In addition, a unique duplication causing a hybrid *PRSS2/PRSS1* gene locus was found in a hereditary pancreatitis family, which further strengthens the clinical importance of CNVs in pancreatitis.[22]

> ***PRSS1* CNVs (duplications) are rare. They typically present as autosomal dominant HP, but other presentations are possible.**

3. PRSS1 exon missense variants.

Some ***PRSS1* exon missense variants** cause *protein misfolding* and can trigger a stress reaction leading to chronic pancreatitis (the best studied examples are *PRSS1* p.R116C, p.C139S, p.L104P, and p.S127C.[23-25] These clinically rare variants are associated primarily with sporadic idiopathic and less frequently with hereditary cases with incomplete penetrance. Phenotypically, the mutant enzymes exhibit reduced or completely diminished secretion, intracellular retention and endoplasmic reticulum (ER) stress.

In addition to *PRSS1*, other highly expressed pancreatic digestive enzymes synthesized in the acinar cell can trigger a similar response. Common examples include *CPA1* variants and the *CEL-HYB1* hybrid allele.[26, 27] Notably, some rare *SPINK1, CTRC* and *PNLIP* variants were also identified with secretion defect indicating misfolding.[28-30] While ER stress was confirmed for *PNLIP* variants, it is unlikely that *CTRC* and *SPINK1* mutations pose significant proteotoxic effects through ER stress due to the relatively low expression levels of these enzymes.

Mechanistically the unfolded protein response (**UPR**) can be activated due to accumulation of unfolded proteins in the rough endoplasmic reticulum (RER). Normally these mutant proteins are removed by special elements of the UPR located in the membrane of the RER, ubiquitinated, and taken to the proteasome for degradation into amino acids and recycled. For the acinar cell, the amount of PRSS1 synthesized will overwhelm the system, triggering the UPR, that activates an immune response (but not necessarily acute pancreatitis). The key enzyme linking secreted proteins, such as PRSS1 to ubiquitin is an E3 ligase called ***UBR1*** (which is also a risk gene for Johanson-Blizzard syndrome.[31]). Ellison et al.[32] showed increased rates of cooccurrence of *PRSS1* variants and *UBR1* variants in patients with CP suggesting that it may be a "conditional" risk factor for CP.

> **A small subset of PRSS1 missense mutations have been linked to CP. They rarely present with antecedent AP. The mechanism may require protein plugs to develop in the RER, resulting in secretion defect of the protein and continuous UPR that drives chronic inflammation. Other commonly expressed genes may cause CP by similar mechanisms.**

4. PRSS1 / PRSS2 loss-of-function AP protective variants

While *PRSS1* and *PRSS2* share 90% of sequence identity, no pancreatitis-associated risk variant cases have been confirmed in *PRSS2*. Some rare variants (p.N84K, p.K98X, p.C171X, p.E209D, p.T230I[33-35] were found in CP patients without known clinical significance. The lack of association can be explained by the fact that *PRSS2* has increased sensitivity to chymotrypsin C (CTRC)-mediated degradation, thus individual mutations are improbable to increase *PRSS2* stability enough to achieve intra pancreatic activation.[36] In contrast, the *PRSS2* p.G191R variant has been found to be underrepresented in the CP versus some control populations[37, 38], but not all [39], highlighting the differences in etiologies and minor allele frequencies between populations and ancestries. *In vitro* studies showed that p.G191R variant exhibited complete loss of trypsin activity thereby supporting the notion that loss of function trypsinogen variants provide protection against pancreatitis.

> **Rare variants in *PRSS2* are associated with protection from AP and CP. Other candidate variants need confirmation such as statistical evidence and functional studies.**

5. PRSS1-PRSS2-TRB locus risk haplotype.

There are 5 trypsin genes / pseudo genes on chromosome 7 within the TRB locus including cationic trypsinogen (*PRSS1*, T4) and anionic trypsinogen (*PRSS2*, T8) that are expressed in the pancreas and three pseudogenes, *PRSS3P1* (T5), *PRSS3P2* (T6), and *TRY7* (T7). In 2012, a SNP within this locus was identified at rs10273639, upstream of the *PRSS1* gene, that was strongly associated with chronic pancreatitis – but not recurrent acute pancreatitis.[14] The mechanism of risk was not explained by this variant, since it was part of a haplotype, and its effect on *PRSS1* expression was only marginal.[14, 40]. Yet, the phenotype was confirmed in subsequent replication studies. [21, 41-46]

The genetic mechanisms of risk are related to a large (9 to 20kb) deletion within the *TRB* locus that includes *PRSS3P2* trip *TRY7* (see Mastromatteo[10]; Figure 8). Long-range sequencing demonstrated that all of the trypsinogen genes are linked, and form two common (and 3 uncommon) haplotypes based on whether they contain 5 trypsinogen genes (55.1%) or the deletion with only 3 trypsinogen genes (44.9%).[10] The high-risk haplotype includes rs3757377_C[10], rs10273639_C[14], a *PRSS1* promoter SNP (rs4726576_C)[47] and two synonymous *PRSS1* variants (rs6666_C and rs6667_C) and spans about 1 million base pairs. Pancreatitis risk is linked to the haplotype with pathogenic effects likely linked to presence of the commonly deleted segment and multiple sequence variants that affect the *TRB* repertoires. [40]

There are multiple statistical and functional lines of evidence informing possible mechanism of high risk. First, eQTL linked to rs6667_C (*PRSS1* p.N246N) includes multiple *TRB* gene variable sequences including *TRBV29-1* that localizes to T-cells in the pancreas and *TRBV28* that localizes to multiple tissues (GTEx).[40] Secondly, the risk haplotype (rs6667_C) has a strong eQTL linked to *PRSS2* expression, with normalized effect size increased by 22% (GTEx).[10, 40] Additional pathogenic effects are also likely.

The large *PRSS1-PRSS2* (*-TRB*) High-Risk haplotype includes 5 trypsin-related genes rather than 3 and alters expression of PRSS1 (slightly), PRSS2 (significantly) and alters the T-Cell response to pancreas injury through altered *TRB* sequence repertoires.

6. PRSS1-PRSS2-TRB protective haplotype.

The *PRSS1-PRSS2-TRB* haplotypes with a large 9 to 20 KB deletions are **protective** for chronic pancreatitis, and possibly acute pancreatitis. One contributing factor may be the *PRSS2* p.Thr8Ile variant (rs62473563 C>T, linked with rs17998863 C>T) that alters the sequence for binding of the signal recognition particle (SRP) targeting the RNA for the ER may result in reduction in PRSS2 synthesis.[10] The eQTL for rs62473563_T reduced *PRSS2* expression by about 70-80% and is moderately linked to the *PRSS1-PRSS2-TRB* deletion (protective) haplotype. This variant is common (MAF; T=0.10 in Europeans, T=0.018 in Africans, T=0.03 in South Asians, and T=0.000 in East Asians) and may contribute to the protective effect of the deletion haplotype. In addition, there may be 5' *PRSS2* promoter transcription sites that are lost with the 20 kb deletion that reduces *PRSS2* expression levels.

Thus, the protective effect of the alternate haplotype includes a different TRB repertoires that the risk haplotype and is associated with reduced anionic trypsinogen expression. Other mechanisms are also possible.

The *PRSS1-PRSS2* (*-TRB*) protective haplotype contains a 9 to 20kb deletion with the loss of two trypsinogen pseudogenes. The overall effects of the protective haplotype may include both a reduction in PRSS2 expression and alternative *TRB* repertoires.

Challenges in genotyping the PRSS1-PRSS2-TRB locus.

The importance of the *PRSS1-PRSS2-TRB* locus deletion is highlighted by the confusion that it created in trying to align whole genome and whole exome sequence data using "shot-gun" sequencing of small transcripts (e.g. 150 bp) that were aligned to earlier reference sequences. GRCh37 was erroneously structured as (T4, T5, T6) and excluding *PRSS2* causing *PRSS2* sequences to be aligned to *PRSS1* templates and erroneously identified as *PRSS1* p.N29I hereditary pancreatitis mutation This was corrected with an update (chr7_gl582971_fix) that included all five genes. Then, in the human reference genome GRCh38, the deletion form has been identified as the primary chromosome 7 sequence with *PRSS1, PRSS3P1* and *PRSS2* (chr7:142746720–142778572) while the *alternative contig* of chromosome 7 includes the 5-gene structure of 20K haplotype (possibly a 10.6 kb tandem duplication) with trypsinogen genes *PRSS1, PRSS3P1, PRSS3P2, TRY7* and *PRSS2* (chr7_KI270803v1_alt:7 49 409-8 01 557).[48] Using the alternative contig for aligning short sequences is very important for correctly aligning the short, 150 bp shot-gun sequences among the highly homologous trypsinogen genes and producing correct genotypes.

V. Annotation of genetic variants related to RAP & CP.

The field of **precision medicine** differs from classic **medical genetics** with a strong focus on multiple common risk variants with conditional effects versus rare pathogenic variants in a single gene that is disease-causing and runs in families. The most striking consequence is that some variants that are very important in RAP and CP (e.g. *CTRC* p.G60G and *PRSS1* p.N246N) are considered "benign" in ClinVar and other genetic resources based on submission of annotation from classic genetic companies based on traditional interpretation of specific variants. Indeed, the problem of multiple interpretation of the same variant's pathogenicity in ClinVar has been recognized, but no solutions are available. [49]

Precision Medicine recognizes that common diseases such as pancreatitis, diabetes, liver diseases, etc. are complex because they require the convergence of multiple genetic, environmental and demographic risk factors that cause the development of disease under known or obscure circumstances. Unlike disease-causing pathogenic mutations, each independent variant is neither necessary nor sufficient to cause disease. Importantly, if multiple risk factors are

necessary, then most of them must be *common*, or sufficient combinations of variants would occur extremely rarely (excluding them from being common diseases). For example, if three genetic risk variables were necessary to cause a disease that was in 0.1% of a population (0.001 or 1 in a thousand), then, on average, each risk factor must have a MAF of 10% (0.10 x 0.10 x 0.10 = 0.001). Thus, common variants are conditionally pathogenic.

Second, as noted above, some important pancreatitis risk variants are synonymous SNPs. The *CTRC* p.G60G (rs497078, c.180C>T) does not change the amino acid, but the variant is linked to multiple other risk variants as part of a risk haplotype. The effect is to reduce expression of chymotrypsin C, a pancreatic zymogen that *degrades trypsin* as a protective mechanism. Thus, reduced *CTRC* expression increases trypsin survival within the pancreas, triggering pancreatitis.

Likewise, *PRSS1* p.N246N (p.Asn246Asn, c.738T>C, rs6667) is a synonymous variant, but, as noted above, the rs6667_C haplotype is linked to the *PRSS1-PRSS2-TRB* risk haplotype for chronic pancreatitis.

Third, precision medicine may utilize polygenic risk scores as an adjusted sum of the risk variants that are associated with common disease. These must be interpreted within the context of the whole patient, including emerging signs and symptoms of disease along with informative biomarkers.

Fourth, precision medicine also considers the effects of environmental factors, ancestry, sex, age and biomarkers in concert with genetic risk factors to clarify the mechanisms of complex diseases and to help guide management. These insights are critical for identifying underlying disease mechanisms and targeting therapies.

A more extensive discussion of this topic is needed to compare and contrast Precision Medicine and Medical Genetics. Issues include the role and need for genetic counselors since risk of having multiple common risk variants has minimal impact on family members, and clinical decision support tools that are central to Precision Medicine but of limited value in Medical Genetics. In summary, although precision medicine utilizes genotype results of many genes and intergenic (regulatory) regions, it is more about medicine than genetics.

VI. Resources for *PRSS1* and *PRSS2* variants.

There are limited publicly available resources for defining the effect of individual genetic variants interacting in complex conditions. For education and research about known *PRSS1* and *PRSS2* the authors recommend the Genetic Risk Factors in Chronic Pancreatitis web site: https://pancreasgenetics.org/

References

1. Whitcomb DC, Gorry MC, Preston RA, Furey W, Sossenheimer MJ, Ulrich CD, et al. Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. Nature Genetics. 1996;14(2):141-5.PMID: 8841182
2. Gorry MC, Gabbaizedeh D, Furey W, Gates LK, Jr., Preston RA, Aston CE, et al. Mutations in the cationic trypsinogen gene are associated with recurrent acute and chronic pancreatitis. Gastroenterology. 1997;113(4):1063-8.PMID: 9322498
3. Dawra R, Sah RP, Dudeja V, Rishi L, Talukdar R, Garg P, et al. Intra-acinar trypsinogen activation mediates early stages of pancreatic injury but not inflammation in mice with acute pancreatitis. Gastroenterology. 2011;141(6):2210-7 e2.PMID: 21875495
4. Gui F, Zhang Y, Wan J, Zhan X, Yao Y, Li Y, et al. Trypsin activity governs increased susceptibility to pancreatitis in mice expressing human PRSS1R122H. J Clin Invest. 2019.PMID: 31550238
5. Geisz A, Sahin-Toth M. A preclinical model of chronic pancreatitis driven by trypsinogen autoactivation. Nature communications. 2018;9(1):5033.PMID: 30487519
6. Jancso Z, Sahin-Toth M. Mutation That Promotes Activation of Trypsinogen Increases Severity of Secretagogue-Induced Pancreatitis in Mice. Gastroenterology. 2020;158(4):1083-94.PMID: 31751559
7. Jancso Z, Sahin-Toth M. Chronic progression of cerulein-induced acute pancreatitis in trypsinogen mutant mice. Pancreatology. 2022;22(2):248-57.PMID: 35063369
8. Demcsak A, Sahin-Toth M. Rate of Autoactivation Determines Pancreatitis Phenotype in Trypsinogen Mutant Mice. Gastroenterology. 2022;163(3):761-3.PMID: 35667407
9. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24.PMID: 25741868
10. Mastromatteo S, Chen A, Gong J, Lin F, Thiruvahindrapuram B, Sung WWL, et al. High-quality read-based phasing of cystic fibrosis cohort informs genetic understanding of disease modification. HGG Adv. 2023;4(1):100156.PMID: 36386424
11. Bryan E, Teague JE, Eligul S, Arkins WC, Moody DB, Clark RA, et al. Human Skin T Cells Express Conserved T-Cell Receptors that Cross-React with Staphylococcal Superantigens and CD1a. J Invest Dermatol. 2024;144(4):833-43 e3.PMID: 37951348
12. Zhong Z, Wu H, Zhang Q, Zhong W, Zhao P. Characteristics of T cell receptor repertoires of patients with acute myocardial infarction through high-throughput sequencing. J Transl Med. 2019;17(1):21.PMID: 30634977

13. Ciacchi L, Farenc C, Dahal-Koirala S, Petersen J, Sollid LM, Reid HH, et al. Structural basis of T cell receptor specificity and cross-reactivity of two HLA-DQ2.5-restricted gluten epitopes in celiac disease. J Biol Chem. 2022;298(3):101619.PMID: 35065967
14. Whitcomb DC, Larusch J, Krasinskas AM, Klei L, Smith JP, Brand RE, et al. Common genetic variants in the CLDN2 and PRSS1-PRSS2 loci alter risk for alcohol-related and sporadic pancreatitis. Nature genetics. 2012;44(12):1349-54.PMID: 23143602
15. Le Marechal C, Masson E, Chen JM, Morel F, Ruszniewski P, Levy P, et al. Hereditary pancreatitis caused by triplication of the trypsinogen locus. Nat Genet. 2006;38(12):1372-4.PMID: 17072318
16. Masson E, Le Marechal C, Chandak GR, Lamoril J, Bezieau S, Mahurkar S, et al. Trypsinogen copy number mutations in patients with idiopathic chronic pancreatitis. Clin Gastroenterol Hepatol. 2008;6(1):82-8.PMID: 18063422
17. Chauvin A, Chen JM, Quemener S, Masson E, Kehrer-Sawatzki H, Ohmle B, et al. Elucidation of the complex structure and origin of the human trypsinogen locus triplication. Hum Mol Genet. 2009;18(19):3605-14.PMID: 19584086
18. Larusch J, Barmada MM, Solomon S, Whitcomb DC. Whole exome sequencing identifies multiple, complex etiologies in an idiopathic hereditary pancreatitis kindred. JOP : Journal of the pancreas. 2012;13(3):258-62.PMID: 22572128
19. Athwal T, Huang W, Mukherjee R, Latawiec D, Chvanov M, Clarke R, et al. Expression of human cationic trypsinogen (PRSS1) in murine acinar cells promotes pancreatitis and apoptotic cell death. Cell death & disease. 2014;5:e1165.PMID: 24722290
20. Wan J, Haddock A, Edenfield B, Ji B, Bi Y. Transgenic expression of human PRSS2 exacerbates pancreatitis in mice. Gut. 2020;69(11):2051-2.PMID: 31974135
21. Zou WB, Cooper DN, Masson E, Pu N, Liao Z, Ferec C, et al. Trypsinogen (PRSS1 and PRSS2) gene dosage correlates with pancreatitis risk across genetic and transgenic studies: a systematic review and re-analysis. Hum Genet. 2022;141(8):1327-38.PMID: 35089416
22. Masson E, Le Marechal C, Delcenserie R, Chen JM, Ferec C. Hereditary pancreatitis caused by a double gain-of-function trypsinogen mutation. Hum Genet. 2008;123(5):521-9.PMID: 18461367
23. Kereszturi E, Szmola R, Kukor Z, Simon P, Weiss FU, Lerch MM, et al. Hereditary pancreatitis caused by mutation-induced misfolding of human cationic trypsinogen: a novel disease mechanism. Hum Mutat. 2009;30(4):575-82.PMID: 19191323
24. Balazs A, Hegyi P, Sahin-Toth M. Pathogenic cellular role of the p.L104P human cationic trypsinogen variant in chronic pancreatitis. Am J Physiol Gastrointest Liver Physiol. 2016:ajpgi 00444 2015.PMID: 26822915
25. Thiel F, Reiser M, Weiss FU. A rare PRSS1 p.S127C mutation is associated with chronic pancreatitis and causes misfolding-induced ER-stress. Pancreatology. 2022;22(8):1112-9.PMID: 36369231
26. Witt H, Beer S, Rosendahl J, Chen JM, Chandak GR, Masamune A, et al. Variants in CPA1 are strongly associated with early onset chronic pancreatitis. Nat Genet. 2013;45(10):1216-20.PMID: 23955596
27. Cassidy BM, Zino S, Fjeld K, Molven A, Lowe ME, Xiao X. Single nucleotide polymorphisms in CEL-HYB1 increase risk for chronic pancreatitis through proteotoxic misfolding. Hum Mutat. 2020;41(11):1967-78.PMID: 32906201
28. Kiraly O, Wartmann T, Sahin-Toth M. Missense mutations in pancreatic secretory trypsin inhibitor (SPINK1) cause intracellular retention and degradation. Gut. 2007;56 Epub .(10):1433-8.PMID: 17525091
29. Beer S, Zhou J, Szabo A, Keiles S, Chandak GR, Witt H, et al. Comprehensive functional analysis of chymotrypsin C (CTRC) variants reveals distinct loss-of-function mechanisms associated with pancreatitis risk. Gut. 2012;62(11):1616-24.PMID: 22942235
30. Szabo A, Xiao X, Haughney M, Spector A, Sahin-Toth M, Lowe ME. A novel mutation in PNLIP causes pancreatic triglyceride lipase deficiency through protein misfolding. Biochim Biophys Acta. 2015;1852(7):1372-9.PMID: 25862608
31. Zenker M, Mayerle J, Lerch MM, Tagariello A, Zerres K, Durie PR, et al. Deficiency of UBR1, a ubiquitin ligase of the N-end rule pathway, causes pancreatic dysfunction, malformations and mental retardation (Johanson-Blizzard syndrome). Nat Genet. 2005;37(12):1345-50.PMID: 16311597
32. Ellison MA, Spagnolo DM, Shelton C, Orlova E, LaRusch J, Whitcomb DC, et al. Complex Genetics in Pancreatitis: Insights Gained From a New Candidate Locus Panel. Pancreas. 2020;49(7):983-98.PMID: 32658084
33. Sofia VM, Da Sacco L, Surace C, Tomaiuolo AC, Genovese S, Grotta S, et al. Extensive molecular analysis suggested the strong genetic heterogeneity of idiopathic chronic pancreatitis. Mol Med. 2016;22.PMID: 27264265
34. Sofia VM, Surace C, Terlizzi V, Da Sacco L, Alghisi F, Angiolillo A, et al. Trans-heterozygosity for mutations enhances the risk of recurrent/chronic pancreatitis in patients with Cystic Fibrosis. Mol Med. 2018;24(1):38.PMID: 30134826
35. Bousman CA, Stevenson JM, Ramsey LB, Sangkuhl K, Hicks JK, Strawn JR, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6, CYP2C19, CYP2B6, SLC6A4, and HTR2A Genotypes and Serotonin Reuptake Inhibitor Antidepressants. Clin Pharmacol Ther. 2023;114(1):51-68.PMID: 37032427
36. Jancso Z, Sahin-Toth M. Tighter Control by Chymotrypsin C (CTRC) Explains Lack of Association between Human

Anionic Trypsinogen and Hereditary Pancreatitis. J Biol Chem. 2016;291(25):12897-905.PMID: 27129265
37. Witt H, Sahin-Toth M, Landt O, Chen JM, Kahne T, Drenth JP, et al. A degradation-sensitive anionic trypsinogen (PRSS2) variant protects against chronic pancreatitis. Nat Genet. 2006;38(6):668-73.PMID: 16699518
38. Masamune A. Genetics of pancreatitis: the 2014 update. Tohoku J Exp Med. 2014;232(2):69-77.PMID: 24522117
39. Mahurkar S, Bhaskar S, Reddy DN, Rao GV, Singh SP, Thomas V, et al. The G191R variant in the PRSS2 gene does not play a role in protection against tropical calcific pancreatitis. Gut. 2009;58(6):881-2.PMID: 19433599
40. Fu D, Blobner BM, Greer PJ, Lafyatis R, Bellin MD, Whitcomb DC, et al. Pancreatitis-Associated PRSS1-PRSS2 Haplotype Alters T-Cell Receptor Beta (TRB) Repertoire More Strongly Than PRSS1 Expression. Gastroenterology. 2023;164(2):289-92 e4.PMID: 36191639
41. Avanthi SU, Ravi Kanth VV, Agarwal J, Lakhtakia S, Gangineni K, Rao GV, et al. Association of claudin2 and PRSS1-PRSS2 polymorphisms with idiopathic recurrent acute and chronic pancreatitis: A case-control study from India. J Gastroenterol Hepatol. 2015;30(12):1796-801.PMID: 26110235
42. Derikx MH, Kovacs P, Scholz M, Masson E, Chen JM, Ruffert C, et al. Polymorphisms at PRSS1-PRSS2 and CLDN2-MORC4 loci associate with alcoholic and non-alcoholic chronic pancreatitis in a European replication study. Gut. 2015;64(9):1426-33.PMID: 25253127
43. Masamune A, Nakano E, Hamada S, Kakuta Y, Kume K, Shimosegawa T. Common variants at PRSS1-PRSS2 and CLDN2-MORC4 loci associate with chronic pancreatitis in Japan. Gut. 2015;64(8):1345-6.PMID: 26002935
44. Paliwal S, Bhaskar S, Nageshwar Reddy D, Rao GV, Thomas V, Singh SP, et al. Association Analysis of PRSS1-PRSS2 and CLDN2-MORC4 Variants in Nonalcoholic Chronic Pancreatitis Using Tropical Calcific Pancreatitis as Model. Pancreas. 2016;45(8):1153-7.PMID: 26784911
45. Rosendahl J, Kirsten H, Hegyi E, Kovacs P, Weiss FU, Laumen H, et al. Genome-wide association study identifies inversion in the CTRB1-CTRB2 locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis. Gut. 2017.PMID: 28754779
46. Weiss FU, Hesselbarth N, Parniczky A, Mosztbacher D, Lammerhirt F, Ruffert C, et al. Common variants in the CLDN2-MORC4 and PRSS1-PRSS2 loci confer susceptibility to acute pancreatitis. Pancreatology. 2018.PMID: 29884332
47. Boulling A, Sato M, Masson E, Genin E, Chen JM, Ferec C. Identification of a functional PRSS1 promoter variant in linkage disequilibrium with the chronic pancreatitis-protecting rs10273639. Gut. 2015;64(11):1837-8.PMID: 26246244
48. Lou H, Xie B, Wang Y, Gao Y, Xu S. Improved NGS variant calling tool for the PRSS1-PRSS2 locus. Gut. 2023;72(1):210-2.PMID: 35288441
49. Lazareva TE, Barbitoff YA, Nasykhova YA, Glotov AS. Major Causes of Conflicting Interpretations of Variant Pathogenicity in Rare Disease: A Systematic Analysis. J Pers Med. 2024;14(8).PMID: 39202055

Corresponding Authors.

David C Whitcomb MD PhD
Pittsburgh, PA
Whitcomb@pancreas.org

And

Andrea Geisz PhD
Boston, MA
geisz@bu.edu

Author Contribution.

DCW conceived of the paper. AG, MW and DCW contributed to drafting the manuscript and approving the final version.

Funding.

There was no funding for this project.

Conflicts of Interest

DCW is a cofounder and consultants to Ariel Precision Medicine., He serves as CSO and Chair, Medical Advisory Board at Ariel Precision Medicine and has equity,

Acknowledgments:

The authors thank Miklos Sahin-Toth MD PhD for helpful discussions.