Title
Global protease activity profiling for pancreatic cancer diagnosis and treatment

Permalink
https://escholarship.org/uc/item/8w55d3jn

Author
Ivry, Sam Lacey

Publication Date
2018

Supplemental Material
https://escholarship.org/uc/item/8w55d3jn#supplemental

Peer reviewed|Thesis/dissertation

Global protease activity profiling for pancreatic cancer diagnosis and
treatment

by

Sam Lacey Ivry

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Acknowledgments

I would like to start off by thanking my Ph.D. advisor, Dr. Charles Craik. As a chemical engineering major, I arrived in the Craik Lab without the typical skill set of most incoming graduate students. Unfazed by this, Charly ensured that I was surrounded by a group of mentors that could provide the training I required and supplement my knowledge where needed. Charly has an amazing ability to bring together fantastic groups of scientists and this really shows in the lab he has built and the collaborations he is continually establishing. His collaborative spirit is inspiring and I hope to carry some of this forward into my career. Charly takes mentorship very seriously and this really shows to his graduate students. Although he is always available to chat, he allows us to explore our own interests and develop into independent scientist. My projects has had many twists and turns throughout the last four years and it is great to have a mentor that is supportive of new avenues that I'm interested in exploring.

I would also like to thank my co-advisor, Dr. Kimberly Kirkwood. Over the last four years, Kim and I have had the opportunity to build new projects, attend conferences, and write and present grants together. I greatly value the guidance she has continually provided me throughout all of these steps. More generally, Kim is just a great person to be around and it has been an absolute pleasure to get to know both her and her family. I look forward to being able to turn to her as a mentor throughout my career.

In addition to Charly, Dr. James Wells and Dr. Kevan Shokat were the other two members of thesis committee. Jim and Kevan were incredibly supportive of both my project and career development goals. Moreover, they never run out of interesting ideas regarding new scientific avenues to explore.

I would next like to thank everyone in the Craik Lab. In particular, Dr. Michael Winter and Dr. Anthony O'Donoghue provided the training I needed to get my project off the ground. Michael is an incredibly rigorous scientist and it was great to learn from someone with such keen attention to detail. Matthew Ravalin has also been a great resource in the lab and has taught me sufficient chemistry to successfully make peptides. Beyond peptide synthesis assistance, Matt is a constant source of interesting ideas for new projects and is one of the most creative scientists I know. The Craik Lab has been a wonderful place – personally and scientifically – for carrying out my doctoral research. Every member is always willing to lend a hand or help talk through experimental roadblocks. I will truly miss spending my lunches with all the lab members in the "interaction room."

I've had fantastic collaborators at UCSF that have all been a pleasure to work with. Dr. Jeremy Sharib has been a great clinical resource and is always available to help put grants and papers together. Dr. Giselle Knudsen has provided endless assistance on all things mass spectrometry. I've also had the pleasure of working with her on more business-oriented projects and her constant enthusiasm makes it easy to get things done. My rotation advisor in the Wells Lab, Alex Martinko, has been another great collaborator. He is an incredibly hard-working scientist, but is somehow always able to find time to help out. I would also like to acknowledge Joseph Lobel. Although we never directly collaborated on a project, he has given me countless pieces of scientific advice. Joe never stops thinking about science and this makes him a great asset to all that work with him – including myself. Many others have contributed to my projects while at UCSF

and include, Dr. Nilotpal Roy, Dr. Grace Kim, Dr. Jonathan Boyce, Dr. Victoria Butler, Dr. Willian Cortopassi, Dr. Arun Wiita, and Dr. Adam Olshen.

The PSPG graduate program has been great to be a part of. My starting class had only five students, Deanna Brackman, Angel Ku, Matthew Chang, and Joel Mefford, but I've really enjoyed working with and getting to know all of them. Our program director, Dr. Deanna Kroetz, has been another great supporter throughout my time in the PSPG program. She cares deeply about the success of each graduate student and will always find time to speak with you regarding any issues that come up. Moving forward, I'm extremely grateful to have been a part of the PSPG program and know that I will continue to rely on the network I've built through this.

Lastly, I would like to thank my family. Although my brother, Henry Ivry, and I have been separated by one continent and international border, he has remained a consistent source of support. My parents, Ann Lacey and Richard Ivry, are my constant champions and it has been a real treat to attend graduate school only 30 minutes away from where they live.

Chapter 1 is a reprint of a review paper published in Protein Science with Nicole O. Meyer, Michael B. Winter, Markus F. Bohn, Giselle M. Knudsen, Anthony J. O'Donoghue, and Charles S. Craik. As first author, I wrote the manuscript and helped design all figures. Chapter 2 is a reprint of work published in Biological Chemistry with Anthony J. O'Donoghue, Chaity Chaudhury, Daniel R. Hostetter, Douglas Hanahan, and Charles S. Craik. My contribution consisted of various biochemical assessments of recombinant procathepsin E. Chapter 3 is a reprint of a first author paper published in Clinical Cancer Research with Jeremy M. Sharib, Dana A. Dominguez, Nilotpal Roy,

Stacy E. Hatcher, Michele T. Yip-Schneider, C. Max Schmidt, Randall E. Brand, Walter G. Park, Matthias Hebrok, Grace E. Kim, Anthony J. O'Donoghue, Kimberly S. Kirkwood, and Charles S. Craik. I designed and conducted all experiments, analyzed data, and wrote the manuscript.

Global protease activity profiling for pancreatic cancer diagnosis and treatment

By Sam Lacey Ivry

Abstract

Proteolytic enzymes are central regulators of many of the hallmarks of cancer development and progression. Understanding how proteolytic activity evolves over the course of tumor development has the potential to provide mechanistic insights into disease progression and enable personalized treatment regimens. In this work, we sought to characterize protease activity in pancreatic ductal adenocarcinoma (PDAC) and its precursor lesions. In order to study proteolysis in a global and unbiased manner, we applied a technology developed in our lab, termed Multiplex Substrate Profiling by Mass Spectrometry (MSP-MS). This technology utilizes a physicochemical diverse peptide library and mass spectrometry to determine protease specificity from various biological samples.

We first used MSP-MS to determine that a PDAC cell line secretes the lysosomal, aspartyl protease, cathepsin E. Secreted cathepsin E primarily exists as a zymogen and displays minimal activity at neutral pH. However, we were able to use MSP-MS to identify two substrates that were cleaved by the cathepsin E zymogen at pH 6.5, suggesting that this enzyme could be playing a proteolytic role in cancer progression. We also used our MSP-MS assay to assess global proteolytic activity in fluid from pancreatic cysts. Mucinous pancreatic cysts are precursor lesions to PDAC and should be resected if high-grade dysplasia or invasive cancer is suspected (HGD/IC). Unfortunately, clinicians are often unable to determine the stage preoperatively or differentiate mucinous from nonmucinous cysts, which are fully benign lesions of the pancreas. Through MSP-MS coupled with proteomic analysis, we identified three acid-activated proteases with increased activity in mucinous pancreatic cysts: cathepsin E, gastricsin, and tripeptidyl

peptidyl peptidase 1 (TPP1). We developed a simple, fluorescence-based assay for analysis of all three proteases and applied this to a cohort of 110 cyst fluid samples. Analysis of gastricsin activity was 95% accurate for differentiating mucinous from nonmucinous cysts, significantly outperforming the most commonly applied biomarker, CEA. Combined analysis of gastricsin with CEA and cathepsin E improved accuracy to 99%. TPP1 activity demonstrated 89% sensitivity and 40% specificity for distinguishing mucinous cysts with HGD/IC from those with only LGD. This performance compares favorably to commonly assessed clinical and radiographic features. Taken together, our results demonstrate that protease activity is differentially regulated in PDAC and activity analysis can serve as an accurate diagnostic tool for helping clinicians identify cysts with the highest potential for malignant transformation.

Table of Contents

List of Tables

List of Figures

Chapter 4

Chapter 1. Global substrate specificity profiling of post-translational modifying enzymes

1.1 Abstract

Enzymes that modify the proteome, referred to as post-translational modifying (PTM) enzymes, are central regulators of cellular signaling. Determining the substrate specificity of PTM enzymes is a critical step in unraveling their biological functions both in normal physiological processes and in disease states. Advances in peptide chemistry over the last century have enabled the rapid generation of peptide libraries for querying substrate recognition by PTM enzymes. In this review, we highlight various peptide-based approaches for analysis of PTM enzyme substrate specificity. We focus on the application of these technologies to proteases but also discuss specific examples in which they have been used to uncover the substrate specificity of other types of PTM enzymes, such as kinases. In particular, we highlight our Multiplex Substrate Profiling by Mass Spectrometry (MSP-MS) assay, which uses a rationally designed, physicochemically diverse library of tetradecapeptides. We show how this method has been applied to PTM enzymes to uncover biological function, as well as guide substrate and inhibitor design. We also briefly discuss how this technique can be combined with other methods to gain a systems-level understanding of PTM enzyme regulation and function.

1.2 Introduction

The primary mechanism by which the diversity of the proteome is increased is through the post-translational modification of proteins. PTM enzymes are responsible for over 200 kinds of modifications of protein substrates and can be divided into two distinct mechanistic categories: *(1)* enzymes that hydrolyze peptide bonds (proteases) and *(2)* enzymes that covalently modify amino acid side chains. PTM enzymes constitute over 5% of the human genome but most have

yet to be fully characterized.[1] An important aspect of understanding the functions of these enzymes requires developing *in vitro* assays in which their specificity and activity can be monitored. Although a variety of assays exist for profiling PTM enzyme specificity, there is particular value in assays in which post-translational modifications of peptide substrates are quantitatively and directly measured. To facilitate this type of assay format, researchers have taken advantage of synthetic peptide chemistry to develop large and diverse peptide libraries.

Peptide synthesis was pioneered by the work of Emil Fischer and Ernest Fourneau who synthesized the dipeptide glycylglycine in 1901. This work laid the foundation for subsequent advances in peptide synthesis with the Nobel Prize in Chemistry being awarded to Bruce Merrifield in 1984 for the development of solid-phase peptide synthesis (SPPS).[2] Merrifield's strategy involved assembly of a peptide chain in a stepwise manner with one end of the nascent peptide anchored to a solid resin until completion of synthesis. Covalent attachment of the growing peptide chain to a solid support renders it insoluble, which facilitates easier transition between synthetic steps, such as washing away excess reactant and byproduct. SPPS has been further streamlined over the last several decades and Fmoc SPPS is currently the most widely used synthetic strategy.[3] Generating synthetic peptides using this technology gained popularity when biologists recognized that synthetic peptides could be used for antibody selection and production.[4] Fmoc SPPS is now easily accomplished using highly automated work-flows.[5,6]

In this review, we first discuss how SPPS has been applied to generate large, highly diverse peptide libraries for the analysis of protease substrate specificity. We provide an overview of the various methods and describe several applications of how these methods have been applied to develop selective protease substrates and inhibitors. We extend this to a discussion of how SPPS has enabled the development of peptide libraries for determining the

specificity of other types of PTM enzymes. Proteome-derived peptide libraries as well as phage and bacterial display have also been widely applied for analysis of protease substrate specificity; however, these technologies are not the focus of the current review and have been reviewed elsewhere.[7–9]

1.3 Peptide-based technologies for analysis of protease specificity

Proteases are one of the largest classes of PTM enzymes, with over 550 encoded in the human genome.[10] These enzymes are essential for normal cellular functions and are implicated in a variety of diseases, such as cancer, neurodegeneration, and blood clotting disorders. Because of the size and importance of this enzyme class, substantial effort has been put into the development of peptide-based technologies for determining protease substrate specificity. Proteases generally recognize substrates in an extended linear conformation, making this class of enzymes particularly amenable to analysis with peptide-based profiling methods.[11]

Traditionally, identification of protease substrates relied on relatively small collections of synthetic peptides with sequences derived from proteins that were known to be proteolyzed. Peptides would be incubated with a target protease and their cleavage assessed, generally through high-performance liquid chromatography (HPLC) with mass spectrometry for cleavage site identification.[12] Once initial substrates were identified, new substrates with variations at select positions would be synthesized to explore subsite specificity. The development of colorimetric and fluorescent peptide substrates simplified cleavage assessment,[13–15] however, defining protease substrate specificity remained an iterative and tedious process. Over the last two decades, this process has been transformed by the development of large and highly diverse peptide libraries. Table 1 summarizes some of the peptide-based technologies for analysis of

protease specificity. The information determined through these approaches can be used for a number of important applications. For example, selective substrates can be designed that enable the real-time monitoring of proteolysis *in vitro* and *in vivo*. Peptide substrates also can be converted into protease inhibitors through coupling to an electrophilic warhead. Furthermore, since proteolytic enzymes recognize their substrates as linear motifs of extended beta strands, specificity information can be used to prioritize potential endogenous substrates. For example, a number of computational approaches have been developed to predict caspase and granzyme B substrates using specificity data determined through peptide-based profiling methodologies.[16,17] Synthetic substrate synthesis, inhibitor design, and endogenous substrate identification are all critical steps in improving our understanding of the biological role of a given protease in both cellular function and pathogenesis.

On-bead fluorescent peptide libraries

In the late 1980s, Kahne *et. al.* developed a radioassay to monitor the hydrolysis of bead-bound peptides.[18] Although this assay was used to assess the half-life of an amide bond in neutral water, which remarkably was determined to be approximately seven years, the authors suggested that this technique could also be used to monitor proteolysis. Meldal *et. al.* were the first to carry that idea forward with the development of an on-bead, combinatorial peptide library for assessing protease specificity.[19] These combinatorial libraries were constructed through split peptide synthesis, leading to a single peptide sequence being present on each bead. The bead-conjugated peptides were constructed with a C-terminal fluorophore and N-terminal quencher, resulting in fluorescence quenching through Förster resonance energy transfer (FRET) prior to release of the fluorophore via proteolytic cleavage.[20,21] Positions C-terminal and N-terminal to the site of

proteolytic cleavage are commonly referred to as prime (P′) and nonprime (P) positions, respectively.[22] Following treatment with a protease of interest, the fluorescent beads were isolated and subjected to Edman degradation where N-terminal residues are labeled and sequentially released to reconstruct the amino acid sequence. However, because each bead contained a mixture of intact and cleaved peptide, it was not possible to differentiate residues released from the native or neo-N-termini generated through proteolysis. This generally meant that the cleavage site needed to be predefined to accurately determine if amino acids detected during Edman degradation were from the intact or cleaved peptide. Once the ratio of intact to cleaved peptide was determined, this was used to calculate percent conversion and estimate catalytic efficiency.[19]

Positional scanning substrate libraries

One of the more widely applied, fluorescence-based techniques for analysis of protease specificity involves the generation of positional scanning-synthetic combinatorial libraries (PS-SCLs). PS-SCLs consist of distinct pools of peptides in which an amino acid in one position is fixed within each pool, while the other positions contain a mixture of amino acids. Initial PS-SCLs were used to determine the P4-P2 specificity of proteases and incorporated a 7-amino-4-methylcoumarin (AMC) fluorophore at the P1′ position (Figure 1A).[23] These libraries were restricted to certain P1 residues, such as aspartic acid and lysine, because the SPPS protocol required attachment of the growing peptide chain to the solid support through the P1 amino acid side chain. The development of a bifunctional 7-amino-4-carbamoylmethylcoumarin (ACC) fluorophore, which can be directly attached to a solid support, enabled the development of PS-SCLs with diversity at the P1 position (Figure 1).[24,25] P1 diversity significantly increased the

number of proteases amenable to analysis through the PS-SCL approach. PS-SCLs have been used to profile the P4-P1 specificity of diverse proteases, including cysteine cathepsins,[26] kallikreins,[27] caspases,[28] and granzymes.[29]

Drag and co-workers recently reported the development of ACC positional scanning libraries incorporating up to 110 unnatural amino acids in the P1 to P4 positions.[30,31] These extended libraries, termed hybrid combinatorial substrate libraries (HyCoSuL), were used to identify a neutrophil elastase substrate with the highest reported catalytic efficiency.[30] The increased chemical space explored through HyCoSuL has also enabled the development of selective caspase substrates.[31] Furthermore, a Counter Selection Substrate Library (CoSeSuL) approach against caspases has been used to develop highly selective legumain probes.[32]

PS-SCLs have also been designed to profile prime side specificity through the incorporation of FRET-based quenching.[33–38] FRET-based PS-SCLs contain a fluorophore and quencher separated by several amino acids. Unlike AMC- and ACC-based PS-SCLs, fluorescence occurs following proteolytic cleavage between any of the amino acids in the peptide substrate. Therefore, mass spectrometry is required to validate the site of cleavage and reconstruct a specificity profile. Recently, Poreba *et. al.* combined FRET and non-FRET based approaches to determine the optimal nonprime and prime side specificity of serine, cysteine and metalloproteases.[39]

Electrophile-based libraries

An alternative design for positional scanning libraries incorporates electrophilic 'warheads' that have been widely used for activity-based profiling of enzymes.[40,41] The electrophile is placed at the P1 position and covalently labels the active-site nucleophile of the target protease.

Electrophile-based libraries contain diversity in the P4-P2 positions and generally use a radiolabel for quantitation of protease labeling.[42,43] These libraries have been successfully used to profile a number of proteases, including cysteine cathepsins, calpains,[42] and the proteasome subunits.[43] Electrophile-based libraries are currently limited to cysteine, serine, and threonine proteases because of the requirement of an active site nucleophile and cannot be used to determine nonprime side specificity. However, a unique advantage of these libraries is that they can be readily converted into protease inhibitors, as specificity information is determined in the context of the electrophilic warhead that can be used in an inhibitor. Although beyond the scope of this review, large libraries of electrophile containing compounds and fragments have been applied for identifying PTM enzyme inhibitors.[44,45]

Microarray peptide libraries

Fluorescent peptide libraries are also commonly used in microarray formats to profile protease substrate specificity.[46,47] Peptide microarrays consist of fluorescent substrates that are spatially separated on a microarray surface either by direct covalent attachment or through individual nanodroplets.[46,48,49] Unlike PS-SCLs, which use pools of fluorescent substrates, cleavage of individual substrate sequences can be directly assessed with microarrays. This enables the highly multiplexed determination of kinetic parameters for each of the typically hundreds of substrates that are evaluated in a given experiment. Moreover, proteases often exhibit subsite cooperativity due to shared determinants of substrate specificity among binding pockets and the optimal positioning of amino acids within the target sequence. Such subsite cooperativity information is lost when using techniques that rely on pools of substrates, but is readily assessable with microarrays because of the spatial separation of individual substrates. One drawback of the

peptide microarrays developed to date is the relatively low sequence diversity as compared to positional scanning methods. With a few notable exceptions,[50] peptide microarrays have been most successful when exploring nonprime specificity and other methods are required in order to query prime side specificity.

Mixture-based oriented peptide libraries

Edman degradation of mixture-based oriented peptide libraries has been used to determine both prime side and nonprime side protease specificity within the same assay format. This profiling strategy developed by Turk *et. al.* uses two separate peptide libraries.[51] First, a fully randomized 12-mer peptide library is synthesized with acetylated N-termini. This library is partially degraded with a protease of interest, releasing C-terminal cleavage products with free N-termini. Edman degradation of the C-terminal cleavage products is used to determine the frequency of each amino acid at each of the prime side positions. A second 12-mer peptide library is then synthesized that incorporates a randomized N-terminal 6-mer and the most favorable P1′ to P6′ amino acids for the C-terminal portion. The predefined prime side sequence directs cleavage to the middle of each peptide in the library. All peptides also contain a C-terminal biotin and free N-termini. Following protease treatment, C-terminal cleavage products are removed with immobilized avidin and the remaining N-terminal products are sequenced by Edman degradation to determine the nonprime side specificity. This approach has been successfully applied to a range of proteases, including matrix metalloproteases,[51,52] anthrax lethal factor,[53] and the serine proteases HtrA1/2.[54,55] Although this is a highly versatile technique, one limitation is that new peptide libraries generally need to be synthesized for each investigated protease. Furthermore, these libraries work best for proteases that have strong prime side specificity determinants, as a

sequence is required to direct cleavage to the middle of each peptide in the second peptide library.

1.4 Multiplex Substrate Profiling by Mass Spectrometry

Mass spectrometry in combination with proteome-derived substrate libraries has been successfully applied to define protease specificity.[56–61] These "degradomics" methods use liquid chromatography with tandem mass spectrometry (LC-MS/MS) to identify the site of proteolytic cleavage within proteome-derived peptides and can define prime side and nonprime side specificity determinants in a single assay. These techniques are quite powerful, but require chemical labeling steps for enrichment and identification of neo-termini and present a challenge in extracting kinetic parameters.

MSP-MS was developed to provide simple, yet highly sensitive and quantitative assay for assessing the extended substrate specificity of proteases. This technique currently uses a library of 228 synthetic tetradecapeptides that contain maximal physicochemical diversity within a minimal sequence space.[62,63] This library was designed based on the observation that most proteases require two optimally positioned amino acids for substrate recognition and cleavage. This phenomenon is generally referred to as the "two-site hypothesis." For example, the specificity of granzyme B is dominated by a preference for isoleucine at the P4 position and aspartic acid at the P1 position.[24,62] As is evident in the crystal structure of granzyme B, two prominent cavities on the enzyme surface (S1 and S4) accommodate these residues and are the primary determinants of substrate recognition (Figure 2).[64] Though these sites are not the only determinants of enzyme efficiency, they contribute to greater than 70% of the binding energy required for substrate recognition and turnover. Numerous proteases appear to follow the two-

site hypothesis with the sites being juxtaposed, for example, on either side of the scissile bond or separated in space along the substrate by one or two amino acids. Therefore, physicochemical diversity in the MSP-MS library was generated through incorporation of all neighbor (XY) and near-neighbor (X*Y, X**Y) amino acid pairings. This simple and chemically defined library enables facile extraction of kinetic parameters for each substrate and is readily amenable to profiling the specificity of purified proteases or complex biological samples without the need for enrichment strategies.

For MSP-MS specificity determination, a recombinant protease or other biological sample of interest is incubated with the peptide library and aliquots are removed at multiple time points (Figure 3). Cleavage sites within library peptides are then identified through LC-MS/MS analysis of each time point and specificity is visualized using a sequence logo, which displays protease amino acid preference relative to the site of cleavage. Label-free quantitation of both parent peptides and their corresponding cleavage products over time can be used to determine kinetic parameters for substrate hydrolysis. This information is critical for the prioritization of optimal sequences for substrate and inhibitor design.

Substrate specificities of a wide range of proteases from all major classes have been interrogated using the MSP-MS assay. Furthermore, because the termini of the peptides are unmodified, the library is well suited for profiling exopeptidase specificity. In particular, the MSP-MS assay has been used to profile carboxypeptidases, such as PRCP,[62] which are generally not amenable to analysis with most of the previously discussed methods that use peptides modified with reporter groups. Furthermore, the MSP-MS assay has been used to identify the prime side specificity determinants of aminopeptidases, such as aminopeptidase N.[65]

The ability of the MSP-MS library to readily profile mixtures of proteases has transformed our ability to quantitatively characterize proteolytic activity in complex biological systems. The assay was recently used to profile the catalytic subunits of the *Plasmodium falciparum* proteasome.[66] The specificity differences between this and the human proteasome were then used to rationally design selective peptidic inhibitors that attenuated malaria development *in vivo*. The MSP-MS assay has been used with protease inhibitors, gene deletions, and immunodepletion in combination with traditional proteomic methods to identify component proteases that are highly active in complex biological samples. This has enabled the "deconvolution" of proteolytic signatures from fungal pathogens,[63,67,68] parasitic organisms,[69,70] cancer cell lines,[62,71] and patient samples,[72,73] and allowed for the prioritization of proteases based on their functional contribution to the global substrate specificity profile. For example, the MSP-MS assay was used to analyze the global activity signatures of the opportunistic fungal pathogen *Candida albicans* in the planktonic and biofilm states.[67] Comparison of the activity signatures from each state coupled to inhibitor and proteomic analysis revealed that two specific secreted aspartyl proteases (Saps) are upregulated during biofilm growth and are critical to biofilm formation *in vitro* and *in vivo*. MSP-MS was also recently used to identify two human aspartyl proteases that are selectively upregulated in cystic precursor lesions of pancreatic cancer.[73] This strategy enabled the development of a highly accurate diagnostic assay using fluorogenic substrates for differentiating benign and premalignant lesions within a cohort of patient cyst fluid samples.

1.5 Kinase specificity analysis using peptide library-based methods

The MSP-MS assay was designed and validated using proteases primarily because they cleave linear peptides and predominately rely on primary amino acid sequence for substrate recognition. Similarly, kinases typically phosphorylate unstructured regions of proteins, and their specificity is strongly dependent upon the amino acid residues surrounding the phosphoacceptor site.[74] Available crystal structures of eukaryotic kinases reveal that many kinases, as with proteases, bind their substrates in an extended, linear conformation.[75,76] Computational efforts using these crystal structures have been able to successfully identify endogenous substrates, highlighting the importance of linear peptide sequences in kinase substrate specificity.[77]

There are numerous peptide-based approaches for profiling the substrate specificity of kinases.[77–82] These methods primarily employ fluorescence, radioactivity, or colorimetry to detect enzyme activity. For kinase-directed PS-SCLs, a series of biotinylated peptides are generated. The central phosphorylatable residue remains fixed and all other positions are varied to query amino acid preferences.[77] These peptide libraries are assayed in a multi-well format with the kinase of interest and radiolabeled ATP. Aliquots of the reaction are then transferred to a streptavidin-coated membrane, and phosphorylation of each peptide substrate is measured via radiography.[83] Quantification of the amount of phosphorylation is then used to determine the specificity of the kinase of interest.

The MSP-MS assay presents a significant improvement upon traditional methods available for profiling kinases. Reporter groups used in other techniques can interfere with kinase-substrate recognition and radioactivity-based methods have costly disposal and present health hazards. The label-free and unbiased design of the MSP-MS library has made it applicable to the analysis of a wide variety of kinases.[84]

Kinases without previously known substrate preference have been profiled using MSP-MS, allowing for the discovery of their key substrate specificity determinants. In addition, the high sensitivity of this assay can allow for profiling of picomolar amounts of kinase.[84] This enables the profiling of a kinase from a single immunoprecipitation experiment, which is highly advantageous if the enzyme cannot be readily expressed and purified. The MSP-MS assay has also been used to obtain kinetic parameters for phosphorylation of individual peptides within the library. This capability has proven useful for analyzing the effect of interacting factors on kinase substrate specificity and catalytic efficiency. For example, MSP-MS was recently used to interrogate the P-TEFb – HIV-1 Tat interaction. P-TEFb, a human kinase that is integral to Tat's transactivation, phosphorylates RNA Polymerase II and two negative elongation factors in order to overcome the stalled RNA Pol II complex, which then allows transcription of the integrated viral genome to continue. There is controversy as to which site within RNA Pol II that P-TEFb phosphorylates, as well as what effect Tat has on specificity and phosphorylation rate.[85–87] MSP-MS analysis of recombinant and immunoprecipitated P-TEFb revealed that P-TEFb phosphorylates serine 5 within the RNA Pol II C-terminal domain. Analysis of P-TEFb with the addition of Tat revealed that Tat selectively increased the catalytic efficiency of P-TEFb toward peptides that most closely resembled RNA Pol II serine 5.[84]

1.6 MSP-MS analysis of additional PTM enzymes

There are approximately 200 types of PTM enzymes, and many of these enzymes have yet to be characterized.[1] In addition to being able to detect proteolytic cleavage and phosphorylation events, the MSP-MS assay could allow for profiling of a variety of other PTMs (Figure 4A). The mass spectrometry-based detection strategy enables high adaptability because most peptide

modifications can be readily identified using existing MS data analysis packages. As described, this technology is particularly well suited for the specificity analysis of PTM enzymes that recognize peptide substrates in an extended, linear conformation. For example, crystal structures of the kinase CDK2/cyclin A and the histone acetyltransferase HAT1 bound to peptide substrates revealed that substrates adopt a linear conformation in the enzyme active site (Figure 4B-C).[88,89] Initial MSP-MS experiments with other types of PTM enzymes have already been carried out. O-GlcNAc transferase (OGT) was profiled using the MSP-MS assay, resulting in "HexNAc" modifications at serine and threonine residues on five library peptides (unpublished data). Protein arginine N-methyltransferase (PRMT1), an arginine-specific histone methyltransferase, was also assayed using the MSP-MS library. PRMT1 methylated a single peptide in the library at an arginine residue adjacent to a glycine. This "RG" motif aligns with the known PRMT1 substrate motif, "RGG."[90] These preliminary results underscore that a simplified library of linear peptides can be used to obtain relevant PTM specificity information.

1.7 Conclusion

Peptide-based technologies have significantly expanded our ability to profile the substrate specificity of proteases and other PTM enzymes. The technologies discussed in this review can provide complementary information when applied together, enabling a more complete understanding of the specificity of a particular enzyme. The specificity information discovered can be used for a number of important applications. As mentioned previously, PTM enzyme specificity has facilitated the design of highly selective chemical probes, such as protease inhibitors or fluorescent substrates.[30,66] These types of probes have been particularly useful in enabling the noninvasive detection of a variety of cancers.[91,92] Specificity information can also

be used to help better define the biological roles of a given PTM enzyme. For example, PTM enzyme specificity has been used to identify the likely site of post-translational modification within endogenous substrates.[16,17,93] This is particularly important when used in combination with degradomic-based methods or other techniques that identify large potential substrate repertoires. More generally, it is increasingly recognized that PTM enzymes, and particularly proteases, regulate biological systems through large, interconnected enzyme-inhibitor networks.[94–96] Determining specificity is a critical aspect of identifying likely interaction partners within the context of these networks. This, in turn, helps to define novel enzymatic cascades, allowing for the development of a systems-level understanding of PTM enzyme biology.

Peptide-based enzyme profiling technologies have progressed significantly since the advent of SPPS and we expect that they will continue to be critical for improving our understanding of the vast repertoire of cellular functions that PTM enzymes regulate.

Acknowledgements

1.8 References

1. Walsh CT, Garneau-Tsodikova S, Gatto GJ (2005) Protein posttranslational modifications: The chemistry of proteome diversifications. Angew. Chemie - Int. Ed. 44:7342–7372.

2. Merrifield RB (1985) Solid Phase Synthesis (Nobel Lecture). Angew. Chemie Int. Ed. English 24:799–810.

3. Behrendt R, White P, Offer J (2016) Advances in Fmoc solid-phase peptide synthesis. J. Pept. Sci. 22:4–27.

4. Walter G (1986) Production and use of antibodies against synthetic peptides. J. Immunol. Methods 88:149–161.

5. Eberle AN, Atherton E, Dryland A, Sheppard RC (1986) Peptide synthesis. Part 9. Solid-phase synthesis of melanin concentrating hormone using a continuous-flow polyamide method. J. Chem. Soc. Perkin Trans. 1:361.

6. Mijalis AJ, Thomas DA, Simon MD, Adamo A, Beaumont R, Jensen KF, Pentelute BL (2017) A fully automated flow-based approach for accelerated peptide synthesis. Nat. Chem. Biol. 13:464–466.

7. López-Otín C, Overall CM (2002) Protease degradomics: a new challenge for proteomics. Nat. Rev. Mol. Cell Biol. 3:509–19.

8. auf dem Keller U, Schilling O (2010) Proteomic techniques and activity-based probes for the system-wide study of proteolysis. Biochimie 92:1705–14.

9. Vizovišek M, Vidmar R, Fonović M, Turk B (2016) Current trends and challenges in proteomic identification of protease substrates. Biochimie 122:77–87.

10. Puente XS, Sánchez LM, Overall CM, López-Otín C (2003) Human and mouse proteases: a comparative genomic approach. Nat. Rev. Genet. 4:544–58.

11. Diamond SL (2007) Methods for mapping protease specificity. Curr. Opin. Chem. Biol. 11:46–51.

12. Stentz FB, Kitabchi AE, Schilling JW, Schronk LR, Seyer JM (1989) Identification of insulin intermediates and sites of cleavage of native insulin by insulin protease from human fibroblasts. J. Biol. Chem. 264:20275–20282.

13. Zimmerman M, Yurewicz E, Patel G (1976) A New Fluorogenic Substrate for Chymotrypsin. Anal. Biochem. 262:258–262.

14. Zimmerman M, Yurewicz E, Patel G (1977) Sensitive Assays for Trypsin, Elastase, and Chymotrypsin Using New Fluorogenic Substrates. Anal. Biochem. 51:47–51.

15. Mares-Guia M, Shaw E, Cohen W (1967) Studies on the Active Center of Trypsin. J. Biol. Chem. 242:5777–5782.

16. Backes C, Kuentzer J, Lenhof H-P, Comtesse N, Meese E (2005) GraBCas: a bioinformatics tool for score-based prediction of Caspase- and Granzyme B-cleavage sites in protein sequences. Nucleic Acids Res. 33:W208-13.

17. Barkan DT, Hostetter DR, Mahrus S, Pieper U, Wells J a, Craik CS, Sali A (2010) Prediction of protease substrates using sequence and structure features. Bioinformatics 26:1714–22.

18. Kahne D, Still WC (1988) Hydrolysis of a peptide bond in neutral water. J. Am. Chem. Soc. 110:7529–7534.

19. Meldal M, Svendsen I, Breddam K, Auzanneau FI (1994) Portion-mixing peptide libraries of quenched fluorogenic substrates for complete subsite mapping of endoprotease specificity. Proc. Natl. Acad. Sci. U. S. A. 91:3314–3318.

20. Rosse G, Kueng E, Page MG, Schauer-Vukasinovic V, Giller T, Lahm HW, Hunziker P, Schlatter D (2000) Rapid identification of substrates for novel proteases using a combinatorial

peptide library. J Comb Chem 2:461–466.

21. Ekici OD, Zhu J, Chung IYW, Paetzel M, Dalbey RE, Pei D (2009) Profiling the substrate specificity of viral protease VP4 by a FRET-based peptide library approach. Biochemistry 48:5753–5759.

22. Schechter I, Berger A (1967) On the Size of the Active Site in Proteases. I. Papain. Biochem. Biophys. Res. Commun. 2:157–162.

23. Rano T a, Timkey T, Peterson EP, Rotonda J, Nicholson DW, Becker JW, Chapman KT, Thornberry N a (1997) A combinatorial approach for determining protease specificities: application to interleukin-1beta converting enzyme (ICE). Chem. Biol. 4:149–155.

24. Harris JL, Backes BJ, Leonetti F, Mahrus S, Ellman J a, Craik CS (2000) Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. Proc. Natl. Acad. Sci. U. S. A. 97:7754–9.

25. Backes BJ, Harris JL, Leonetti F, Craik CS, Ellman J a (2000) Synthesis of positional-scanning libraries of fluorogenic peptide substrates to define the extended substrate specificity of plasmin and thrombin. Nat. Biotechnol. 18:187–193.

26. Choe Y, Leonetti F, Greenbaum DC, Lecaille F, Bogyo M, Brömme D, Ellman JA, Craik CS (2006) Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. J. Biol. Chem. 281:12824–12832.

27. Debela M, Magdolen V, Schechter N, Valachova M, Lottspeich F, Craik CS, Choe Y, Bode W, Goettig P (2006) Specificity profiling of seven human tissue kallikreins reveals individual subsite preferences. J. Biol. Chem. 281:25678–88.

28. Relationships F, For E, Mediators KEY, Apoptosis OF, Thornberry NA, Rano TA, Peterson EP, Rasper DM, Timkey T, Garcia-calvo M, et al. (1997) A Combinatorial Approach Defines

Specificities of Members of the Caspase Family and Granzyme B. 272:17907–17911.

29. Mahrus S, Craik CS (2005) Selective chemical functional probes of granzymes A and B reveal granzyme B is a major effector of natural killer cell-mediated lysis of target cells. Chem. Biol. 12:567–77.

30. Kasperkiewicz P, Poreba M, Snipas SJ, Parker H, Winterbourn CC, Salvesen GS, Drag M (2014) Design of ultrasensitive probes for human neutrophil elastase through hybrid combinatorial substrate library profiling. Proc. Natl. Acad. Sci. U. S. A. 111:2518–23.

31. Poreba M, Kasperkiewicz P, Snipas SJ, Fasci D, Salvesen GS, Drag M (2014) Unnatural amino acids increase sensitivity and provide for the design of highly selective caspase substrates. Cell Death Differ. 21:1482–92.

32. Poreba M, Solberg R, Rut W, Lunde NN, Kasperkiewicz P, Snipas SJ, Mihelic M, Turk D, Turk B, Salvesen GS, et al. (2016) Counter Selection Substrate Library Strategy for Developing Specific Protease Substrates and Probes. Cell Chem. Biol. 23:1023–1034.

33. Cotrin SS, Puzer L, De Souza Judice WA, Juliano L, Carmona AK, Juliano MA (2004) Positional-scanning combinatorial libraries of fluorescence resonance energy transfer peptides to define substrate specificity of carboxydipeptidases: Assays with human cathepsin B. Anal. Biochem. 335:244–252.

34. Bersanetti PA, Andrade MCC, Casarini DE, Juliano MA, Nchinda AT, Sturrock ED, Juliano L, Carmona AK (2004) Positional-scanning combinatorial libraries of fluorescence resonance energy transfer peptides for defining substrate specificity of the angiotensin I-converting enzyme and development of selective C-domain substrates. Biochemistry 43:15729–15736.

35. Tanskul S, Oda K, Oyama H, Noparatnaraporn N, Tsunemi M, Takada K (2003) Substrate specificity of alkaline serine proteinase isolated from photosynthetic bacterium, Rubrivivax

gelatinosus KDDS1. Biochem. Biophys. Res. Commun. 309:547–551.

36. Puzer L, Cotrin SS, Alves MFM, Egborge T, Araújo MS, Juliano MA, Juliano L, Brömme D, Carmona AK (2004) Comparative substrate specificity analysis of recombinant human cathepsin V and cathepsin L. Arch. Biochem. Biophys. 430:274–283.

37. Wysocka M, Lesner A, Majkowska G, Łęgowska A, Guzow K, Rolka K, Wiczk W (2010) The new fluorogenic substrates of neutrophil proteinase 3 optimized in prime site region. Anal. Biochem. 399:196–201.

38. Wysocka M, Wojtysiak A, Okonska M, Gruba N, Jarzab M, Wenta T, Lipinska B, Grzywa R, Sienczyk M, Rolka K, et al. (2015) Design and synthesis of new substrates of HtrA2 protease. Anal. Biochem. 475:44–52.

39. Poreba M, Szalek A, Rut W, Kasperkiewicz P, Rutkowska-Wlodarczyk I, Snipas SJ, Itoh Y, Turk D, Turk B, Overall CM, et al. (2017) Highly sensitive and adaptable fluorescence-quenched pair discloses the substrate specificity profiles in diverse protease families. Sci. Rep. 7:43135.

40. Sanman LE, Bogyo M (2014) Activity-based profiling of proteases. Annu. Rev. Biochem. 83:249–73.

41. Niphakis MJ, Cravatt BF (2014) Enzyme inhibitor discovery by activity-based protein profiling. Annu. Rev. Biochem. 83:341–77.

42. Greenbaum DC, Arnold WD, Lu F, Hayrapetian L, Baruch A, Krumrine J, Toba S, Chehade K, Brömme D, Kuntz ID, et al. (2002) Small molecule affinity fingerprinting: A tool for enzyme family subclassification, target identification, and inhibitor design. Chem. Biol. 9:1085–1094.

43. Nazif T, Bogyo M (2001) Global analysis of proteasomal substrate specificity using positional-scanning libraries of covalent inhibitors. Proc. Natl. Acad. Sci. 98:2967–2972.

44. Bachovchin DA, Koblan LW, Wu W, Liu Y, Li Y, Zhao P, Woznica I, Shu Y, Lai JH,

Poplawski SE, et al. (2014) A high-throughput, multiplexed assay for superfamily-wide profiling of enzyme activity. Nat. Chem. Biol. 10:656–663.

45. Backus KM, Correia BE, Lum KM, Forli S, Horning BD, González-Páez GE, Chatterjee S, Lanning BR, Teijaro JR, Olson AJ, et al. (2016) Proteome-wide covalent ligand discovery in native biological systems. Nature 534:570–574.

46. Salisbury CM, Maly DJ, Ellman JA (2002) Peptide microarrays for the determination of protease substrate specificity. J. Am. Chem. Soc. 124:14868–14870.

47. Winssinger N, Damoiseaux R, Tully D, Geierstanger B, Burdick K, Harris JL (2004) PNA-Encoded Protease Substrate Microarrays. Chem. Biol. 128:189–190.

48. Gosalia DN, Salisbury CM, Ellman J a, Diamond SL (2005) High throughput substrate specificity profiling of serine and cysteine proteases using solution-phase fluorogenic peptide microarrays. Mol. Cell. Proteomics 4:626–36.

49. Gosalia DN, Salisbury CM, Maly DJ, Ellman JA, Diamond SL (2005) Profiling serine protease substrate specificity with solution phase fluorogenic peptide microarrays. Proteomics 5:1292–1298.

50. Barrios AM, Craik CS (2002) Scanning the prime-Site substrate specificity of proteolytic enzymes: A novel assay based on ligand-Enhanced lanthanide ion fluorescence. Bioorganic Med. Chem. Lett. 12:3619–3623.

51. Turk BE, Huang LL, Piro ET, Cantley LC (2001) Determination of protease cleavage site motifs using mixture-based oriented peptide libraries. Nat. Biotechnol. 19:661–7.

52. Park HI, Turk BE, Gerkema FE, Cantley LC, Sang QXA (2002) Peptide substrate specificities and protein cleavage sites of human endometase/matrilysin-2/matrix metalloproteinase-26. J. Biol. Chem. 277:35168–35175.

53. Turk BE, Wong TY, Schwarzenbacher R, Jarrell ET, Leppla SH, Collier RJ, Liddington RC, Cantley LC (2004) The structural basis for substrate and inhibitor selectivity of the anthrax lethal factor. Nat. Struct. Mol. Biol. 11:60–66.

54. Martins LM, Turk BE, Cowling V, Borg A, Jarrell ET, Cantley LC, Downward J (2003) Binding specificity and regulation of the serine protease and PDZ domains of HtrA2/Omi. J. Biol. Chem. 278:49417–49427.

55. Chien J, He X, Shridhar V (2009) Identification of tubulins as substrates of serine protease HtrA1 by mixture-based oriented peptide library screening. J. Cell. Biochem. 107:253–263.

56. Schilling O, Overall CM (2008) Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. Nat. Biotechnol. 26:685–94.

57. Biniossek ML, Niemer M, Maksimchuk K, Mayer B, Fuchs J, Huesgen PF, McCafferty DG, Turk B, Fritz G, Mayer J, et al. (2016) Identification of Protease Specificity by Combining Proteome-Derived Peptide Libraries and Quantitative Proteomics. Mol. Cell. Proteomics 15:mcp.O115.056671.

58. Staes A, Van Damme P, Helsens K, Demol H, Vandekerckhove J, Gevaert K (2008) Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). Proteomics 8:1362–1370.

59. Kleifeld O, Doucet A, auf dem Keller U, Prudova A, Schilling O, Kainthan RK, Starr AE, Foster LJ, Kizhakkedathu JN, Overall CM (2010) Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. Nat. Biotechnol. 28:281–8.

60. Vizovišek M, Vidmar R, Van Quickelberghe E, Impens F, Andjelković U, Sobotič B, Stoka V, Gevaert K, Turk B, Fonović M (2015) Fast profiling of protease specificity reveals similar

substrate specificities for cathepsins K, L and S. Proteomics 15:1–12.

61. Mahrus S, Trinidad JC, Barkan DT, Sali A, Burlingame AL, Wells J a (2008) Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. Cell 134:866–76.

62. O'Donoghue AJ, Eroy-reveles AA, Knudsen GM, Ingram J, Zhou M, Statnekov JB, Greninger AL, Hostetter DR, Qu G, Maltby DA, et al. (2012) Global identification of peptidase specificity by multiplex substrate profiling. Nat. Methods 9: 1095-1100.

63. Knudsen GM, Beekman C, Perry JA, Johnson AD, Derisi JL, Craik CS, O'Donoghue AJ (2015) Destructin-1 is a collagen-degrading endopeptidase secreted by Pseudogymnoascus destructans, the causative agent of white-nose syndrome. Proc. Natl. Acad. Sci. U. S. A. 112:E3152.

64. Waugh SM, Harris JL, Fletterick R, Craik CS (2000) The structure of the pro-apoptotic protease granzyme B reveals the molecular determinants of its specificity. Nat. Struct. Biol. 7:762–765.

65. Joshi S, Chen L, Winter MB, Lin Y-L, Yang Y, Shapovalova M, Smith PM, Liu C, Li F, LeBeau AM (2017) The Rational Design of Therapeutic Peptides for Aminopeptidase N using a Substrate-Based Approach. Sci. Rep. 7:1424.

66. Li H, O'Donoghue AJ, van der Linden WA, Xie SC, Yoo E, Foe IT, Tilley L, Craik CS, da Fonseca PCA, Bogyo M (2016) Structure- and function-based design of Plasmodium-selective proteasome inhibitors. Nature 530:233–236.

67. Winter MB, Salcedo EC, Lohse MB, Hartooni N, Gulati M, Sanchez H, Takagi J, Hube B, Andes DR, Johnson AD, et al. (2016) Global Identification of Biofilm-Specific Proteolysis in Candida albicans. MBio 7:1–13.

68. Clarke SC, Dumesic PA, Homer CM, O'Donoghue AJ, La Greca F, Pallova L, Majer P, Madhani HD, Craik CS (2016) Integrated Activity and Genetic Profiling of Secreted Peptidases in Cryptococcus neoformans Reveals an Aspartyl Peptidase Required for Low pH Survival and Virulence. PLoS Pathog. 12.

69. Goupil LS, Ivry SL, Hsieh I, Suzuki BM, Craik CS, O'Donoghue AJ, McKerrow JH (2016) Cysteine and Aspartyl Proteases Contribute to Protein Digestion in the Gut of Freshwater Planaria. PLoS Negl. Trop. Dis. 10:e0004893.

70. Dvořák J, Fajtová P, Ulrychová L, Leontovyč A, Rojo-Arreola L, Suzuki BM, Horn M, Mareš M, Craik CS, Caffrey CR, et al. (2015) Excretion/secretion products from Schistosoma mansoni adults, eggs and schistosomula have unique peptidase specificity profiles. Biochimie :1–11.

71. Donoghue AJO, Ivry SL, Chaudhury C, Hostetter DR, Hanahan D, Craik CS (2016) Procathepsin E is highly abundant but minimally active in pancreatic ductal adenocarcinoma tumors. Biol. Chem. 397:871–881.

72. O'Donoghue AJ, Jin Y, Knudsen GM, Perera NC, Jenne DE, Murphy JE, Craik CS, Hermiston TW (2013) Global substrate profiling of proteases in human neutrophil extracellular traps reveals consensus motif predominantly contributed by elastase. PLoS One 8:e75141.

73. Ivry SL, Sharib JM, Dominguez DA, Roy N, Hatcher SE, Yip-Schneider M, Schmidt CM, Brand RE, Park WG, Hebrok M, et al. (2017) Global protease activity profiling provides differential diagnosis of pancreatic cysts. Clin. Cancer Res. 23:4865–4874.

74. Songyang Z, Lu KP, Kwon YT, Tsai L-H, Filhol O, Cochet C, Brickey DA, Soderling TR, Bartleson C, Graves DJ, et al. (1996) A Structural Basis for Substrate Specificities of Protein Ser/Thr Kinases: Primary Sequence Preference of Casein Kinases I and II, NIMA, Phosphorylase

Kinase, Calmodulin- Dependent Kinase II, CDK5, and Erk1. Mol. Cell. Biol. 16:6486–6493.

75. Brinkworth RI, Breinl RA, Kobe B (2003) Structural basis and prediction of substrate specificity in protein serine/threonine kinases. Proc. Natl. Acad. Sci. U. S. A. 100:74–79.

76. Johnson LN, Lowe ED, Noble MEM, Owen DJ The structural basis for substrate recognition and control by protein kinases. In: FEBS Letters. Vol. 430. ; 1998. pp. 1–11.

77. Hutti JE, Jarrell ET, Chang JD, Abbott DW, Storz P, Toker A, Cantley LC, Turk BE (2004) A rapid method for determining protein kinase phosphorylation specificity. Nat. Methods 1:27–29.

78. Wu JJ, Phan H, Lam KS (1998) Comparison of the intrinsic kinase activity and substrate specificity of c-Abl and Bcr-Abl. Bioorganic Med. Chem. Lett. 8:2279–2284.

79. Rychlewski L, Kschischo M, Dong L, Schutkowski M, Reimer U (2004) Target Specificity Analysis of the Abl Kinase using Peptide Microarray Data. J. Mol. Biol. 336:307–311.

80. Rodriguez M, Li SSC, Harper JW, Songyang Z (2004) An Oriented Peptide Array Library (OPAL) Strategy to Study Protem-Protein Interactions. J. Biol. Chem. 279:8802–8807.

81. Luo K, Zhou P, Lodish HF (1995) The specificity of the transforming growth factor beta receptor kinases determined by a spatially addressable peptide library. Proc. Natl. Acad. Sci. U. S. A. 92:11761–11765.

82. Pearson RB, Kemp BE (1991) Protein kinase phosphorylation site sequences and consensus specificity motifs: tabulations. Methods Enzymol. 200:62–81.

83. Chen C, Turk BE (2010) Analysis of serine-threonine kinase specificity using arrayed positional scanning peptide libraries. Curr. Protoc. Mol. Biol. Chapter 18.

84. Meyer NO, O'Donoghue AJ, Schulze-Gahmen U, Ravalin M, Moss SM, Winter MB, Knudsen GM, Craik CS (2017) Multiplex Substrate Profiling by Mass Spectrometry for Kinases

as a Method for Revealing Quantitative Substrate Motifs. Anal. Chem. 89:4550–4558.

85. Schüller R, Forné I, Straub T, Schreieck A, Texier Y, Shah N, Decker T-M, Cramer P, Imhof A, Eick D (2016) Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Mol. Cell 61:305–314.

86. He N, Liu M, Hsu J, Xue Y, Chou S, Burlingame A, Krogan NJ, Alber T, Zhou Q (2010) HIV-1 Tat and Host AFF4 Recruit Two Transcription Elongation Factors into a Bifunctional Complex for Coordinated Activation of HIV-1 Transcription. Mol. Cell 38:428–438.

87. Czudnochowski N, Bösken CA, Geyer M (2012) Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. Nat. Commun. 3:842.

88. Alterio V, De Simone G, Monti SM, Scozzafava A, Supuran CT (2007) Carbonic anhydrase inhibitors: Inhibition of human, bacterial, and archaeal isozymes with benzene-1,3-disulfonamides-Solution and crystallographic studies. Bioorganic Med. Chem. Lett. 17:4201–4207.

89. Cook A, Lowe ED, Chrysina ED, Skamnaki VT, Oikonomakos NG, Johnson LN (2002) Structural studies on phospho-CDK2/cyclin A bound to nitrate, a transition state analogue: Implications for the protein kinase mechanism. Biochemistry 41:7301–7311.

90. Wooderchak WL, Zang T, Zhou ZS, Acuña M, Tahara SM, Hevel JM (2008) Substrate profiling of PRMT1 reveals amino acid sequences that extend beyond the "RGG" paradigm. Biochemistry 47:9456–9466.

91. Blum G, von Degenfeld G, Merchant MJ, Blau HM, Bogyo M (2007) Noninvasive optical imaging of cysteine protease activity using fluorescently quenched activity-based probes. Nat. Chem. Biol. 3:668–677.

92. Whitley MJ, Cardona DM, Lazarides AL, Spasojevic I, Ferrer JM, Cahill J, Lee C-L, Snuderl

M, Blazer DG, Hwang ES, et al. (2016) A mouse-human phase 1 co-clinical trial of a protease-activated fluorescent probe for imaging cancer. Sci. Transl. Med. 8:320ra4.

93. Song R, Oren D a, Franco D, Seaman MS, Ho DD (2013) Strategic addition of an N-linked glycan to a monoclonal antibody improves its HIV-1-neutralizing activity. Nat. Biotechnol. 31:1047–1052.

94. Overall CM, Dean RA (2006) Degradomics: Systems biology of the protease web. Pleiotropic roles of MMPs in cancer. Cancer Metastasis Rev. 25:69–75.

95. Fortelny N, Cox JH, Kappelhoff R, Starr AE, Lange PF, Pavlidis P, Overall CM (2014) Network analyses reveal pervasive functional regulation between proteases in the human protease web. PLoS Biol. 12:e1001869.

96. Mason SD, Joyce J a (2011) Proteolytic networks in cancer. Trends Cell Biol. 21:228–37.

1.9 Figures and tables

A
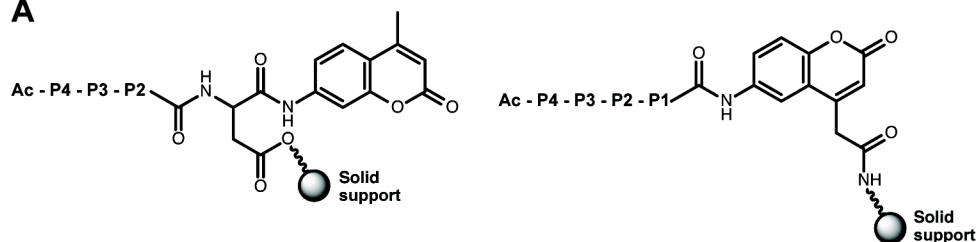
B
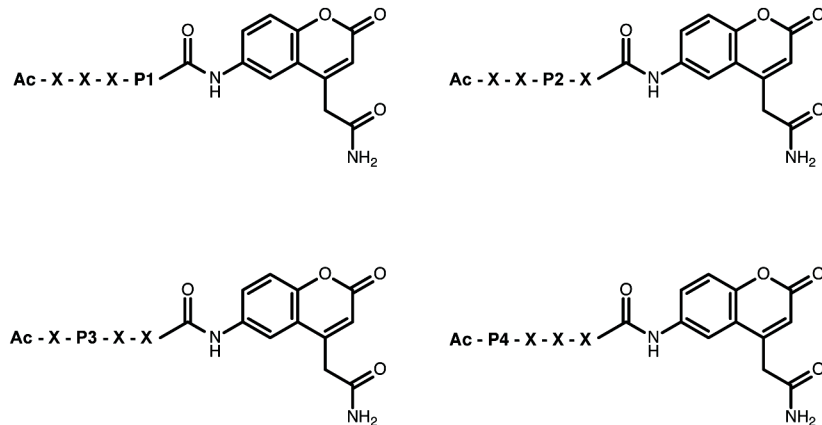
Figure 1.1. Construction of positional scanning-synthetic combinatorial libraries for analysis of prime side specificity. (A) During solid-phase peptide synthesis, 7-amino-4-methylcoumarin (AMC) based positional scanning-synthetic combinatorial libraries are generally conjugated to the solid support via the side chain of the P1 amino acid, while 7-amino-4-carbamoylmethylcoumarin (ACC) based libraries are conjugated directly through the fluorophore. **(B)** The four P_n sublibraries each contain 20 distinct pools of substrates, where one amino acid is fixed at P_n position and the remaining positions (X) contain an equimolar mixture of amino acids.
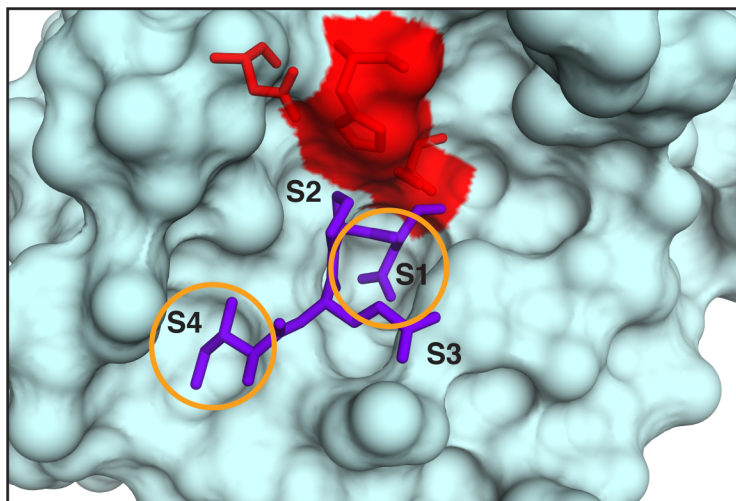
Figure 1.2. The ecotin peptide (purple) binds to the active site of granzyme B in a linear conformation.[64] The catalytic triad is shown in red. Recognition of the peptide, IEPD (written P4-P1), is dominated by the S4 and S1 pockets (circled in yellow).
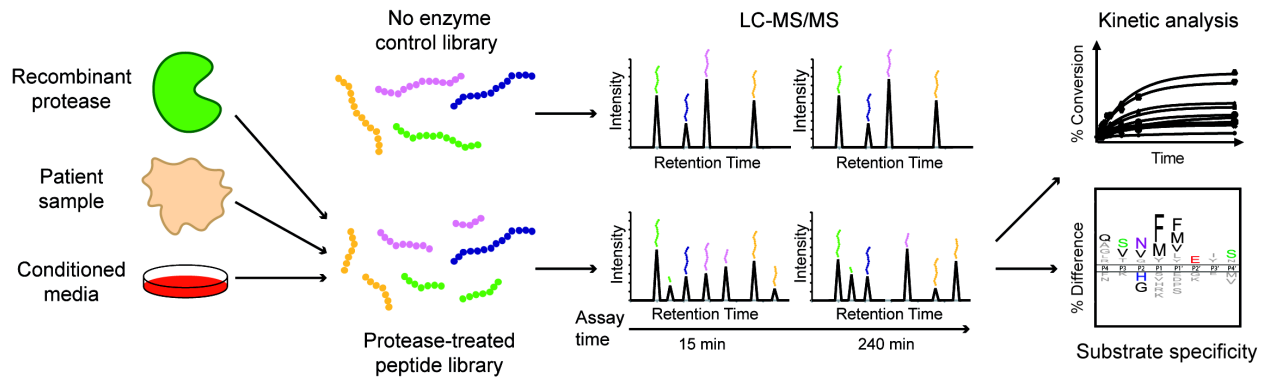
Figure 1.3. MSP-MS workflow for protease specificity determination. Either a recombinant protease, patient sample, conditioned media or other complex, protease-containing biological sample is added to the MSP-MS peptide library. Aliquots are removed at specific time points and peptide cleavage is assessed through LC-MS/MS analysis. Cleavage-site identification can be used to construct a sequence logo representation of the global substrate specificity. Cleavage product quantification enables the kinetic analysis of individual substrate cleavage events.
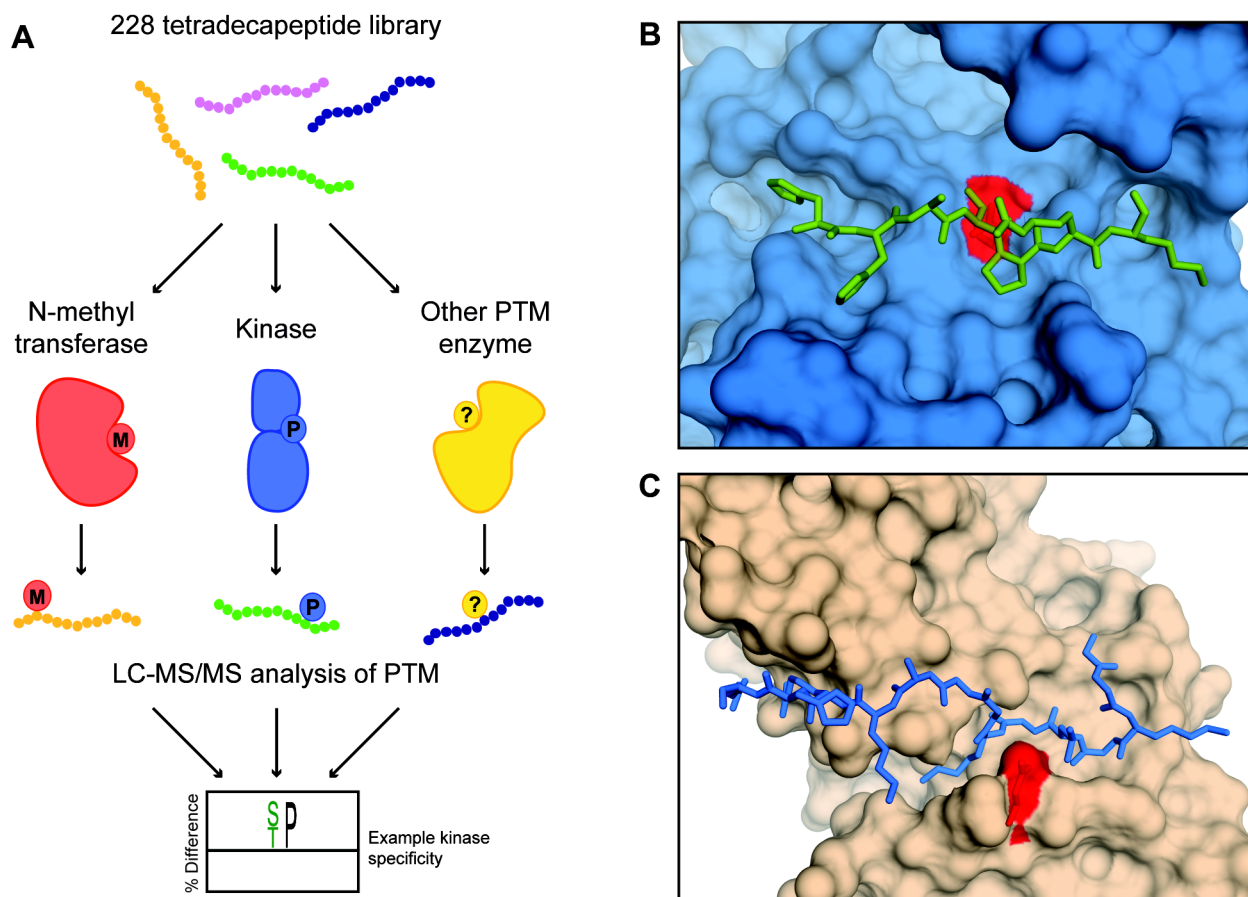
Figure 1.4. Application of the MSP-MS library to other PTM enzymes. (A) A general scheme for the specificity analysis of PTM enzymes using the MSP-MS assay. PTM enzymes are incubated with the peptide library and modification is detected through LC-MS/MS analysis. **(B)** The kinase CDK2/cyclin A (shown in blue) recognizes a peptide substrate (depicted as green sticks) in a linear conformation.[89] The active site lysine residue is colored red. **(C)** Similarly, the histone acetyltransferase HAT1 also binds a linear substrate.[88]

31

Tables

Table 1.1. Peptide-based protease activity profiling technologies.

Peptide-based profiling technology	Advantages	Disadvantages
On-bead FRET libraries	High sequence diversity. Useful for analysis of subsite cooperativity.	Requires prior knowledge of specificity. On-bead immobilization can produce artifacts. New libraries need to be synthesized for each protease.
Positional scanning-synthetic combinatorial libraries	High sequence diversity, especially for newer libraries incorporating unnatural amino acids. Simple and validated in-solution fluorescent assay.	Generally limited to nonprime specificity profiling. Does not provide information related to subsite cooperativity.
Electrophile-based libraries	High sequence diversity. Already contain electrophilic warhead for conversion into protease inhibitor.	Generally limited to nonprime specificity profiling. Primarily limited to serine, threonine, and cysteine proteases.
Peptide microarrays	Can be used to determine kinetic parameters for substrate hydrolysis. Able to analyze protease activity in complex biological samples. Useful for analysis of subsite cooperativity.	Limited sequence diversity. Immobilization of peptide substrates can produce artifacts. Limited to nonprime specificity profiling.
Mixture-based oriented peptide libraries	High sequence diversity. Can be used to profile prime and nonprime specificity.	New libraries often need to be synthesized for nonprime side profiling. Primarily limited to proteases with strong prime side specificity. Requires enrichment step.
Multiplex substrate profiling by mass spectrometry	Can profile prime and nonprime specificity. Can be used to determine kinetic parameters for hydrolysis of substrates. Useful for analysis of subsite cooperativity. Global activity in complex samples can be analyzed. Rationally designed peptide library.	Relativity limited sequence diversity. Activity is not monitored in real-time.

Chapter 2. Procathepsin E is highly abundant but minimally active in pancreatic ductal adenocarcinoma tumors

2.1 Abstract

Pancreatic ductal adenocarcinoma (PDAC) is the fourth leading cause of cancer related death in the United States. To improve patient outcomes for this lethal cancer, there has been a significant effort to improve our understanding of the molecular processes underlying malignant progression. The cathepsin family of lysosomal proteases is being increasingly recognized for their altered expression in PDAC and role in facilitating cancer progression. The aspartyl protease cathepsin E is overexpressed in a number of cancers and has been investigated as a biomarker for PDAC. Using a mass spectrometry based substrate-profiling assay, we previously identified cathepsin E as the dominant protease activity in conditioned media from a mouse PDAC cell line (O'Donoghue et al., 2012). In this study, we show that this PDAC cell line overexpresses cathepsin E over 400-fold and secretes the protease as a zymogen. By decreasing the pH, we induce removal of the cathepsin E prodomain and activate proteolysis. In a PDAC genetic model, we detect cathepsin E in precursor lesions and observe protein accumulation over the course of disease progression with cathepsin E accounting for more than 3% of the total protein in mice with end-stage disease. Cathepsin E again only exists in its proform and treatment of PDAC mice with a cathepsin E inhibitor was unable to decrease tumor burden. Lastly, we used our multiplex substrate profiling by mass spectrometry (MSP-MS) assay to identify two peptides that are hydrolyzed by procathepsin E at pH 6.5 and quantified cleavage by HPLC. This work represents a comprehensive biochemical analysis of cathepsin E proteolytic

activity in PDAC and could facilitate the development of improved biomarkers for disease detection.

2.2 Introduction

With a mean survival rate of six months and five-year survival rate of less than 5%, PDAC remains one of the most lethal cancers (Costello et al., 2012). PDAC can be distinguished from other pancreatic and non-pancreatic malignancies by a characteristic set of mutations, including activation of the oncogene *Kras*, which occurs in 95% of cases (Hezel et al., 2006). *Kras* mutation is thought to initiate the formation of pre-invasive ductal lesions, known as pancreatic intraepithelial neoplasias (PanINs) (Morris et al., 2010). Successive mutations in the tumor suppressor genes *Ink4a* (90%), *Trp53* (75%), and *Smad4* (50%) cause PanINs to undergo graded histological progression and eventual transformation into PDAC (Hezel et al., 2006). Generation of mice harboring these signature genetic mutations has yielded models that closely recapitulate the histopathogenesis of the human disease.

In cancer, dysregulation of protease activity can lead to degradation of the extracellular matrix and facilitate neoplastic progression (Mason and Joyce, 2011). Many studies have focused on the roles of matrix metalloproteases (MMPs) and serine proteases due to their localization on the exterior of the cell (Kessenbrock et al., 2010; Sevenich and Joyce, 2014). In PDAC, silencing of the metalloprotease ADAM17 markedly reduced invasiveness and migration of cancer cells (Ringel et al., 2006). Cysteine proteases of the papain subfamily, known as cysteine cathepsins, are being increasingly investigated for their role in cancer. These proteases are predominantly found within endolysosomal vesicles, but are upregulated and secreted by cancer cells and thus may play an intracellular and extracellular role in tumor progression (Mohamed and Sloane,

2006). Using a cysteine cathepsin inhibitor, Joyce and colleagues observed defects in tumor growth, invasion, and angiogenesis in a mouse model of pancreatic islet cell carcinoma (Joyce et al., 2004). This phenotype was not observed following treatment with a broad spectrum MMP inhibitor (Bergers et al., 1999). Further studies by the same group determined that deletion of cathepsins B, L, or S in this mouse model correlated with a reduction in tumor burden and invasion (Gocheva et al., 2006, 2010).

Two catalytically distinct members of the cathepsin family are the aspartyl proteases, cathepsins D and E. Cathepsin D is a ubiquitously expressed lysosomal protease (Reid et al., 1989). The proform of the enzyme is overexpressed and secreted by a number of cancer types (Beaujouin et al., 2010; Laurent-Matha et al., 2001). Secreted procathepsin D binds the cell surface and stimulates growth of breast, prostate and lung cancer cells *in vitro* and *in vivo*.

Cathepsin E is an intracellular aspartyl protease found mainly in cells of the immune system, such as lymphocytes (Sakai et al., 1989), microglia (Nishioku et al., 2002), dendritic (Chain et al., 2005), and activated B cells (Burster et al., 2008). Unlike the highly homologous cathepsin D, its intracellular localization varies with cell type. Cathepsin E has been reported to reside in endosomes, the plasma membrane, the endoplasmic reticulum, and the Golgi apparatus (Sakai et al., 1989; Zaidi and Kalbacher, 2008). While the exact physiological role of cathepsin E has yet to be elucidated, some studies indicate that it plays a role in antigen processing via the MHC class II pathway. Cathepsin E knockout mice develop atopic dermatitis-like skin lesions with increased susceptibility to bacterial infection and accumulate lysosomal membrane sialoglycoproteins that result in a novel form of lysosomal storage disorder (Tsukuba et al., 2003, 2006). Overexpression of cathepsin E has been found in gastric, cervical, lung, intestinal and pancreatic cancer (Cruz-Monserrate et al., 2012; Mota et al., 1997; Ullmann et al., 2004).

Cathepsin E detection in the urine of mice with intestinal adenomas was reported as a potential marker for disease progression, while strong expression of cathepsin E in tumors of patients with lung carcinomas correlates with increased survival (Ullmann et al., 2004). In the pancreas, cathepsin E is detectable in early PanIN lesions and accumulates as cells progress to PDAC (Buchholz et al., 2005). While detection of the protein in pancreatic juice of patients has been shown to be a promising diagnostic marker to distinguish PDAC from chronic pancreatitis, no functional studies have been performed to characterize the role of cathepsin E in cancer progression (Uno et al., 2000).

Our group previously identified multiple secreted proteases, including cathepsin E, from a mouse PDAC cell line (O'Donoghue et al., 2012). In addition, we determined that cathepsin E was the dominant proteolytic activity in conditioned media from this cell line when assayed at pH 5.2. Since cathepsin E is generally found intracellularly and is optimally active between pH 3.5 and pH 4.5, we decided to investigate if this enzyme could function in the pericellular space near neutral pH, namely pH 6.5 (Yasuda et al., 1999). We performed a comprehensive biochemical analysis of cathepsin E activation in both a mouse PDAC cell line and in tumors from a PDAC genetic model. Cell line and tumor associated cathepsin E exclusively exists as a proenzyme and tumor growth was not slowed by treatment with an active site inhibitor. However, procathepsin E retains residual proteolytic activity as detected by our MSP-MS assay.

2.3 Results

Only procathepsin E exists in primary PDAC cells

In a previous study, we detected complement factor B, carboxypeptidase E, and cathepsins E, B, and L in conditioned media from a primary PDAC cell line derived from a

mouse tumor (O'Donoghue et al., 2012). When this media was assayed at pH 5.2, cathepsin E activity was responsible for the majority of detected proteolysis. These studies were performed with a global and unbiased substrate-profiling assay that uses mass spectrometry to detect the proteolytic degradation of a synthetic peptide library (O'Donoghue et al., 2012). To uncover the functional role of this enzyme in PDAC, an in-depth biochemical analysis of PDAC derived cathepsin E was required.

In this study, we used the PDAC cell line from our previous work. This cell line is derived from *p48-Cre; Kras*^{G12D}*; Trp53*$^{f/f}$ mice, which develop PDAC that histologically mirrors the human disease (Nolan-Stevaux et al., 2009). Immunoblotting analysis confirmed that cathepsin E was present in conditioned media from this cell line (Figure 1A). As cathepsin E is generally found intracellularly, we first wanted to confirm that the protein found in the media was not simply the result of cellular lysis. To test this, we treated cells with brefeldin A, an inhibitor of the secretory pathway, and confirmed that cathepsin E was no longer present in the conditioned media. Interestingly, with an apparent molecular mass of 53 kDa, the secreted cathepsin E was larger than the 46 kDa protein that was previously observed in a mouse study of atopic dermatitis (Tsukuba et al., 2003). In addition, the molecular mass of the intracellular protein was lower than that of the extracellular protein, indicating that these enzymes were differentially post-translationally modified (Figure 1B). Mouse cathepsin E is synthesized as a 397 amino acid protein, consisting of a 20 amino acid signal peptide, a 39 amino acid propeptide, and a 338 amino acid catalytic domain (Zaidi and Kalbacher, 2008). Cathepsin E also has two N-linked glycosylation sites at asparagines 91 and 323. Like other aspartyl proteases, procathepsin E can auto-activate under acidic conditions resulting in the irreversible hydrolysis of the propeptide (Richter et al., 1998). To determine if the higher molecular weight extracellular

cathepsin E corresponded to the proform, conditioned media was exposed to acidic conditions. This resulted in conversion to a lower molecular weight protein of approximately 49 kDa (Figure 1B). Surprisingly, intracellular cathepsin E at 50 kDa was also converted to a lower molecular weight form (46 kDa) following acid exposure, indicating that both intracellular and extracellular proteins exist in the proform. The intracellular 46 kDa protein is likely the same cathepsin E that was detected by Tsukuba and coworkers (Tsukuba et al., 2003). Treatment of procathepsin E with the deglycosylase enzymes, PNGaseF, resulted in a protein of the same molecular weight, indicating that the mass difference was due to alternative glycosylation.

Sastradipura and coworkers have previously shown that procathepsin E is expressed in rat microglia cells and is rapidly converted to the mature enzyme (Sastradipura et al., 2002). Immunoelectron microscopy of these microglia cells showed that the enzyme was associated with dense endosome-like organelles. Since only procathepsin E was detected in PDAC cellular lysates, we wanted to investigate if intracellular localization might be restricting acid-mediated protease activation. Using immunofluorescence, procathepsin E was detected throughout the cytoplasm and colocalized with early endosomal antigen 1 (EEA1) (Figure 1C). Late endosomes have a pH less than 6.0 and autoactivation of procathepsin E to the mature form occurs between pH 3.0 and pH 6.0 (Cappiello et al., 2004; Rao-Naik et al., 1995). Therefore it is unlikely that procathepsin E proceeds into late endosomes since the mature enzyme is not detected in the whole cell lysate. Instead, procathepsin E remains in early endosomes or is trafficked outside of the cell.

Acidification of tumor lysates activates cathepsin E

We proceeded to further analyze cathepsin E in pancreatic tumors from *p48-Cre; Kras*[G12D]; *Trp53*[f/f] mice. Tumors isolated from 13-14 week old mice were lysed under non-denaturing conditions and incubated at neutral or acidic pH for 15 minutes. Immunoblot analysis could only detect 50 kDa procathepsin E at neutral pH. Exposure to acid caused a shift in molecular weight to 46 kDa, which corresponds to removal of the cathepsin E prodomain and generation of the mature enzyme (Figure 2A). No mature cathepsin E was detected in non-acid treated tumor lysates.

As procathepsin E can be activated to the mature enzyme, we used an internally quenched fluorescent substrate to quantify the activity in tumor lysates. This substrate is selective for cathepsin E over cathepsin D and other acid-acting proteases (Yoshiyuki et al., 2005). At pH 3.5, recombinant mouse cathepsin E rapidly hydrolyzed the internally quenched substrate, while no activity was detected at pH 7.4 (Figure 2B). At neutral pH, low levels of proteolytic activity were detected in tumor extracts from 10 week old PDAC mice and from exocrine pancreatic tissue isolated from non-tumor bearing littermates. This low-level activity is likely due to non-specific cleavage by proteases that are optimally active at neutral pH. When these same protein extracts were assayed at pH 3.5, a 13- to 37-fold increase in activity was detected in the tumor lysates, but not in the pancreatic extracts from healthy littermate controls. It was unclear from these results if cathepsin E activity in tumor extracts was due to the lack of an endogenous protease inhibitor or an increase in cathepsin E expression. To address this concern, procathepsin E protein levels were assessed in pancreatic tissue lysates from PDAC mice and non-tumor bearing littermates.

Procathepsin E is overexpressed in PDAC tumors

Procathepsin E was detected in all tumor extracts by immunoblot, but not in tissue from healthy littermates (Figure 2C). This suggests that malignant transformation in this mouse model is driving enhanced expression of this protein. Interestingly, immunoblot analysis failed to detect procathepsin E or mature cathepsin E in tumor lysate from the RIP1-Tag2 model of pancreatic islet carcinoma. The RIP1-Tag2 mouse model expresses the SV40 T antigen oncogenes in insulin-producing β cells and a number of prior studies have demonstrated that protease overexpression promotes tumorigenesis in this model (Folkman et al., 1989). In fact, the cysteine cathepsins H, L, C, Z, B and S genes were found to be upregulated during RIP1-Tag2 tumorigenesis and these enzymes contributed to angiogenic switching, tumor vascularity, and proliferation (Joyce and Hanahan, 2004; Joyce et al., 2004). Using the same quantitative RT-PCR approach as outlined in the RIP1-Tag2 mouse study, we compared cathepsin expression levels in PDAC tumors to healthy pancreatic tissue. Cathepsin E mRNA expression increased 441-fold in tumors relative to healthy pancreatic tissue, while cathepsins B, H, S, and Z mRNA levels increased by 3-fold or less (Figure 2D). Cathepsin L, C and D expression levels were lower in tumors than in healthy pancreatic tissue. The high expression of cathepsin E relative to the other cathepsins prompted us to further investigate the role of this protease in promoting tumorigenesis in the *p48-Cre; Kras^{G12D}; Trp53^{f/f}* PDAC mouse model. More generally, this points to specific cathepsins playing unique roles in the development of different types of pancreatic tumors.

Cathepsin E expression increases with progressive dysplasia

As procathepsin E was present in PDAC tumors and not in healthy tissue, we investigated the time-dependent expression and localization of this protein in the pancreas of *p48-Cre;*

Kras^{G12D}; Trp53^{f/f} mice as they progress from harboring low-grade PanINs to invasive PDAC. Immunohistochemical staining of the exocrine pancreas of non-tumor bearing mice showed only basal levels of procathepsin E. However, by 4 weeks procathepsin E expression increased and the protein localized to neoplastic ductal structures. Staining intensity increased with progressive stages of dysplasia from 4 weeks to end-stage (Figure 3A-D). Mice succumb to the disease between 13-15 weeks and dense staining is evident in tumors from deceased mice. This unusually strong staining prompted us to quantity the amount of procathepsin E protein in PDAC tumor lysates using a semi-quantitative immunoblot. Surprisingly, approximately 6.25 nanograms of procathepsin E were present in 0.2 µg of total protein extracted from PDAC tumor lysates (Figure 3E). Therefore, in end-stage mouse tumors, procathepsin E accounted for approximately 3% of the total soluble protein extracted from the tumor.

Peptide inhibitors bind to procathepsin E

Structural studies of aspartyl protease zymogens revealed that the propeptide occupies the enzyme active site. Upon acid exposure, structural rearrangements lead to cleavage and dissociation of the propeptide (Ostermann et al., 2004). At pH 7.5 we did not observe procathepsin E activity against a fluorescent substrate (Figure 2B), suggesting that the propeptide is restricting access to the protease active site. We were curious to determine if an inhibitor of cathepsin E could compete with the propeptide for binding to the active site. Pepstatin is a potent inhibitor of cathepsin E and other pepsin-type aspartyl proteases (Dunn, 2002). We have successfully used pepstatin-agarose beads to enrich for aspartyl proteases from complex protein mixtures (Bibo-Verdugo et al., 2015). Using a similar approach, we added pepstatin-agarose beads to tumor lysate and then diluted the mixture in either pH 7.5 lysis buffer or pH 3.5

activation buffer. After 30 minutes incubation at pH 7.5 we were unable to recover procathepsin E. However, incubation at pH 3.5 in the presence of excess pepstatin-agarose led to recovery of procathepsin E and not the lower molecular weight mature protease (Figure 4A). This shows that the cathepsin E prodomain can be displaced prior to cleavage occurring, enabling access to the active site of the proenzyme.

To target procathepsin E in PDAC, we used the FDA approved protease inhibitor, ritonavir. Ritonavir was approved to treat HIV and potently inhibits the viral aspartyl protease, however, it has nanomolar affinity towards cathepsin E (Kempf et al., 1998). Although it is likely that ritonavir has weaker affinity for procathepsin E we decided to treat *Pdx-1-Cre; LSLKras*[G12D]*; Trp53*[f/f] mice with 125 mg/kg of for 28 days. After this period, tumors were removed and weighed. No significant difference in tumor burden was observed between ritonavir and vehicle treated animals (Figure 4B). There are a number of possibilities for why ritonavir failed to reduce tumor volume. PDAC tumors may be refractory to therapeutic intervention using this compound. Alternatively, procathepsin E may not play a critical enzymatic role in tumor progression and the use of a catalytic inhibitor will, therefore, not impact tumor progression.

Procathepsin E retains minimal activity at pH 6.5

Cruz-Monserrate and colleagues have shown that the PDAC tumor microenvironment is acidic and therefore it is possible that conditions exist within the tumor that facilitate procathepsin E activity, but are insufficient to promote auto-activation to the mature enzyme (Cruz-Monserrate et al., 2014). Our studies with pepstatin-agarose showed that the cathepsin E active site is accessible prior to prodomain cleavage. We were unable to detect cleavage using a standard cathepsin E fluorescent substrate, but other substrates may be able to compete with the

cathepsin E propeptide and allow us to detect activity. Therefore, we investigated procathepsin E activity and specificity using the highly sensitive MSP-MS assay. This assay uses tandem mass spectrometry to monitor proteolytic cleavage of 228 synthetic tetradecapeptides. At pH 5.5, procathepsin E was converted to the mature enzyme and 65 unique cleavages were detected following 15 minutes of incubation (Figure 5A, Supplementary Figure 1). The substrate specificity profile under these conditions was similar to what has been previously reported for cathepsin E (Impens et al., 2010; O'Donoghue et al., 2012). When we performed the same assay at pH 6.5, auto-activation of procathepsin E did not occur (Supplementary Figure 1). However, cleavage of two peptides – PHWQRVIFFRLNTP and KWLIHPTFSYnRWP – in the library was evident after 4 hours incubation (Figure 5B). The same peptides were also hydrolyzed by mature cathepsin E at pH 5.5 within 15 minutes.

The mass spectrometry based assay is largely qualitative and therefore we decided to use reversed-phase high performance liquid chromatography (RP-HPLC) to quantify the rate of hydrolysis. One of the peptide substrates is cleaved into two products that each contains a tryptophan residue, allowing for quantitation via tryptophan fluorescence during RP-HPLC analysis. At pH 5.5, 2 nM of mature cathepsin E completely cleaved 100 μM of substrate within 1 hour (Supplementary Figure 2). The two cleavage products eluted from the C18 column between 15.25 and 16.25 minutes. Analysis of the mass of the cleavage products by MALDI-MS demonstrated that cathepsin E cut at the predicted site (Supplementary Figure 3). At pH 6.5, time-dependent product formation was evident using 200 nM of procathepsin E (Figure 5C). After 216 hours incubation, only 2.6% percent of the parent peptide was hydrolyzed.

2.4 Discussion

Previous work by our group determined that a PDAC cell line isolated from a *Pdx-1-Cre; LSLKras*^{G12D}*; Trp53*$^{f/f}$ mouse secretes several endo-lysosomal proteases, such as cathepsin B, L, and E (O'Donoghue et al., 2012). Several studies have looked at the role of the cysteine cathepsins B and L in pancreatitis and pancreatic cancer (Gocheva et al., 2006; Lyo et al., 2012). In fact, these enzymes have been implicated in playing multiple roles in various cancers (Joyce and Hanahan, 2004; Sevenich et al., 2014). However, few studies have focused on the role of the aspartyl cathepsins in cancer, with the notable exception of cathepsin D in breast cancer (Masson et al., 2010). We were particularly intrigued to detect high levels of cathepsin E in conditioned media from PDAC cells and sought to investigate further the expression, localization, and enzymatic function of this enzyme. However, in this study we show that only procathepsin E is secreted from the primary cells in culture and activity is only detectable after the media becomes sufficiently acidic to induce protease auto-activation.

A previous immunohistochemical study by Buchholz and colleagues demonstrated that human cathepsin E is absent from the normal pancreatic duct, but abundant in PDAC and precursor lesions (Buchholz et al., 2005). In agreement with this, we show that cathepsin E in the *Pdx-1-Cre; LSLKras*^{G12D}*; Trp53*$^{f/f}$ mouse model is hyperexpressed in PDAC tumors and accumulates in high abundance in the cytoplasm. In fact, it is likely to be one of the most abundant proteins in the mouse tumor extract and accounts for approximately 3% of the total soluble protein. This intracellular protein exists only as procathepsin E but can be activated upon brief exposure to acid. The absence of mature cathepsin E in the tumor lysate indicates that the protein does not traffic through the normal endo-lysosomal pathway with exposure to a sufficiently acidic environment to cause activation. Therefore, the accumulation of excess intracellular procathepsin E may indicate endo-lysosomal dysfunction in these cells.

The excess procathepsin E identified in PDAC is analogous to procathepsin D secreted by breast cancer cells (Laurent-Matha et al., 2001). Procathepsin D is secreted at low levels by normal mammary epithelial cells, but secretion increases by up to 40-fold in breast cancer cell lines. Several studies have shown that procathepsin D binds to prosaposin in the ER and these proteins are co-secreted (Gopalakrishnan et al., 2004; Laurent-Matha et al., 2002). In addition, procathepsin D can bind to cystatin-C and will degrade this protein upon activation to the mature enzyme at pH 3.5 (Laurent-Matha et al., 2012; Khalkhali-Ellis and Hendrix, 2015). Degradation of this macromolecular inhibitor ultimately results in an increase in cysteine protease activity, which promotes tumor progression and metastasis. A recent study showed that secretion of procathepsin D from the MCF-7 breast cancer cell line is increased under hypoxia (Achour et al., 2016). Hypoxic tumors cause acidification of the tumor microenvironment and Cruz-Monserrate and colleagues demonstrated that an acidic microenvironment clearly exists within PDAC tumors (Cruz-Monserrate et al., 2014). Interestingly, the microenvironment conditions appear to favor the functioning of acid-acting proteases such as those found in the endosomes and lysosomes. However, it appears to be insufficiently acidic to activate procathepsin E in pancreatic tumors and procathepsin D in breast tumors.

While procathepsin E may mediate its functions through binding partners, we decided to investigate if this protein was catalytically active under pH conditions that are relevant to the tumor. We show that the aspartyl protease inhibitor, pepstatin, can compete with the propeptide sequence for binding to the enzyme active site. This prompted us to treat mice with the HIV protease inhibitor ritonavir; however, no reduction in tumor burden was evident. While ritonavir is a potent inhibitor of the mature enzyme (Kempf et al., 1998), it is likely to be much less potent against the zymogen. Therefore, it was unclear if the ineffectiveness of the inhibitor was the

result of cathepsin E activity not being important for tumor progression or if the inhibitor was unable to sufficiently engage with the enzyme.

Although it is well established that protease zymogens can be catalytically active, this has not been demonstrated for procathepsin E. Previous studies have shown that rat cathepsin E alters its substrate specificity at neutral pH and becomes more "trypsin-like", cleaving on the C-terminal side of arginine residues. However, Zaida and coworkers predicted that rat cathepsin E was likely to be contaminated with trace amounts of a "trypsin-like" proteases that are optimally active at neutral pH (Zaidi et al., 2007). Here we used mass spectrometry and a library of 228 tetradecapeptides to detect proteolytic activity of procathepsin E (O'Donoghue et al., 2012). This highly sensitive assay identified two peptide substrates for procathepsin E that are also hydrolyzed by the mature enzyme. The cleavage site within these peptides mirrors the known specificity of cathepsin E. Taken together, these results show that procathepsin E has weak proteolytic activity at pH 6.5 and suggests that the highly abundant zymogen may play an enzymatic role in tumor progression.

While our studies could not detect mature cathepsin E in whole tumor extracts, it is possible that a fraction of the 53 kDa procathepsin E gets converted to the 49 kDa mature form, but is undetectable by immunoblot. A number of imaging agents targeting cathepsin E activity have recently been developed to monitor PDAC status *in vivo* (Cruz-Monserrate et al., 2012; Keliher et al., 2013; Li et al., 2014). These imaging agents were designed to target the mature enzyme active site or be specifically cleaved by cathepsin E. In addition, Abd-Elgaliel and colleagues have developed a cathepsin E activated prodrug that is toxic to cells expressing cathepsin E (Abd-Elgaliel et al., 2013). From our work it is clear that in PDAC, cathepsin E primarily exists as a zymogen with minimal enzymatic activity. It is likely that developing

probes to target procathepsin E may therefore be more effective. For example, radiolabeled antibodies have been successfully used to image proteases that are overexpressed in tumors and could be developed against procathepsin E for non-invasive detection of PDAC (LeBeau et al., 2013).

In conclusion, our analysis of cathepsin E activation in PDAC revealed that the protease is highly overexpressed and exists exclusively in its proform. Cathepsin E may play a role in PDAC progression that is independent of proteolytic activity. Cathepsin E is now recognized as an exciting biomarker for PDAC and this information may facilitate the development of novel imaging agents for monitoring disease status.

2.5 Materials and Methods

Mouse strains, tissue culture, and ritonavir administration

The *p48-Cre; Kras^{G12D}; Trp53$^{f/f}$* mouse strain was used and cells were isolated from PDAC tumors as previously described (Nolan-Stevaux et al., 2009). Cells were maintained in complete DMEM with 10% FBS. Mice were treated for 28 days with 125 mg/kg ritonavir or ethanol by oral gavage. The mouse pancreas was removed and weighed and tumor burden was assessed using the ratio of tumor weight to body weight. All animal studies were conducted in compliance with University of California Institutional Animal Care and Use Committee guidelines.

Histological analysis, immunochemistry, and immunofluorescence

Pancreatic tissue from *p48-Cre; Kras^{G12D}; Trp53$^{f/f}$* mice was harvested after 4 and 10 weeks and from animals with end-stage disease. Tissue was fixed overnight in zinc-buffered

formalin and embedded in paraffin. 5 mm thick sections were subjected to either H&E staining or an antigen retrieval procedure (Citra; BioGenex). Following inhibition of endogenous peroxidases and blocking the slides, goat anti-mouse cathepsin E antibody (1:100; R&D Systems AF1130) was applied overnight at 4°C. Biotinylated or fluorochrome-conjugated antibodies were used as secondary antibodies (1:200; Jackson Immunoresearch). 3-39-DAB tetrahydrochloride (Sigma D4293) was used as a chromogen.

Cathepsin E isolation and western blotting

Mouse tumors and cell lines were lysed in 20mM Tris-HCl pH 7.5, 150mM NaCl, 1% Triton X-100 buffer. Protein lysates were diluted to 1.5 mg/ml in lysis buffer or 1.25 mM sodium acetate pH 3.5 and incubated on ice for 30 minutes. Where appropriate, 20% (v/v) pepstatin-agarose (Sigma) was added to protein lysates and incubated on ice for 30 minutes. Agarose beads were washed in lysis buffer or sodium acetate buffer and protein was eluted in 1X LDS sample buffer containing 20 mM TCEP solution (ThermoFisher Scientific). Secreted cathepsin E was isolated by immunoprecipitation with 2.5 µg of goat anti-mouse cathepsin E antibody (R&D Systems AF1130) attached to Protein G Dynabeads. All samples were subjected to electrophoresis on denaturing 10% NuPAGE Bis-Tris gels (Invitrogen). Proteins were transferred to polyvinylidene fluoride membranes and blocked in Tris buffered saline Triton X-100 (TBST) containing 5% (w/v) milk. Membranes were incubated with a polyclonal rabbit anti-mouse cathepsin E antibody (1:1,000; Thermo Scientific PA3-16821), washed and incubated with goat-derived HRP-conjugated secondary antibody (BioRad). Immunoblots were developed on film with the ECL Plus detection system (GE Healthcare). To verify that equal amounts of protein were being compared across samples, actin levels were quantified in parallel with an

anti-β-actin antibody (1:10,000; Sigma). Where indicated, samples were treated with the deglycosylase PNGaseF (New England Biolabs) according to the manufacturers protocol prior to gel electrophoreses.

Fluorescent protease assays

All assays were performed at room temperature in either PBS containing 0.1% Triton X-100 or 50 mM sodium acetate, pH 3.5, 100 mM NaCl, 0.1% Triton X-100 using 40 μM substrate (Bachem M-2625). Assays were performed in triplicate in black round-bottom 96-well plates on a SpectraMax Gemini fluorescence spectrometer (Molecular Devices) using excitation and emission wavelengths of 328 nm and 393 nm, respectively.

Quantitative PCR

Total RNA was prepared from tumors and cell lines using the RNeasy Mini (Qiagen) according to the manufacturer's recommendations. Total RNA from microdissected samples was prepared using the RNeasy Micro (Qiagen). DNase treatment and RNA cleanup were performed with the DNA-Free RNA Kit (Zymo Research). cDNA synthesis was performed using iScript (Bio-Rad). PCRs were performed using the following TaqMan assays (Applied Biosystems): *Ctsb, Ctsc, Ctsh, Ctsl, Ctss, Ctsx, Ctsd,* and *Ctse.* The *mGus* assay was obtained from Integrated DNA Technologies (F) CTCATCTGGAATTTCGCCGA; (R) GGCGAG TGAAGATCCCCTTC; (Probe) fam-CGAACCAGTCACCGCTGAGAGTAATCG-bhq1 was used to normalize expression. Quantitative PCR reactions were performed on an ABI7900HT Sequence Detection System. Ct values were determined and subtracted to obtain the ΔCt [ΔCt = Ct(test locus) - Ct(control locus)]. Relative fold difference was calculated as $100*2^{-\Delta Ct}$.

Multiplex peptide cleavage assay

Recombinant mouse cathepsin E proteolytic activity was analyzed using the MSP-MS assay as described previously (O'Donoghue et al., 2012). For all assays, an expanded 228 peptide library was used and split into two pools containing 114 peptides at 500 nM each. Assays were performed at pH 6.5 for procathepsin E and pH 5.5 for mature cathepsin E. To promote proteolytic activity of procathepsin E, 250 nM of enzyme was incubated with peptide pools at 37 °C. Mature cathepsin E was assayed at 50 nM. Aliquots were removed and acid-quenched to pH 2 or less with formic acid after 15, 60, 240, and 1200 minutes. To avoid acid-mediated activation of cathepsin E, 10 μM of the aspartyl protease inhibitor pepstatin A (Sigma) was added to assays directly before the acid-quench. Control samples without recombinant cathepsin E were prepared under identical conditions to account for non-enzymatic degradation of the substrates. Prior to LC-MS/MS peptide sequencing, samples were desalted using C18 ZipTips (Millipore).

For LC-MS/MS, an LTQ Orbitrap XL mass spectrometer (Thermo) equipped with a 10,000 psi system nanoACQUITY UPLC instrument (Waters) was used for reversed phase chromatography with a C18 column (1.7 μm bead size, 100 μm x 100 mm). The LC was operated at 600 nL/min flow rate and peptides were separated using a linear gradient over 65 minutes from 2% B to 30% B, with solvent A being 0.1% formic acid in water and solvent B being 0.1% formic acid in 70% acetonitrile. Survey scans were recorded over 350-1800 m/z range and MS/MS was performed with CID fragmentation on the six most intense precursor ions. Mass spectrometry peak lists were generated using in-house software called PAVA. Peak lists were searched in Protein Prospector v. 5.10.0 against a database containing the 228 peptide sequences from the MSP-MS library. For database searching, peptide sequences were allowed to

contain the following variable modifications: oxidation of tryptophan, proline, and phenylalanine and N-terminal pyroglutamic acid from glutamine or glutamic acid. Cleavage site data was extracted from Protein Prospector using the MSP-extractor software. iceLogo software was used to visualize conserved patterns in amino acid sequence at ±4 positions adjacent to the identified site of cleavage (Colaert et al., 2009).

Analysis of cathepsin E activity by HPLC

Activity assays with recombinant cathepsin E and individual substrates were performed in pH 6.5 PBS and pH 5.5 sodium acetate buffer. For pH 6.5 assays, 250 nM of cathepsin E was incubated with 100 μM substrate at 37 °C. After 4, 8, 24, 72, and 216 hours, protease activity was heat inactivated. For pH 5.5 assays, 2 nM of cathepsin E was incubated with 100 μM substrate at 37 °C. Cathepsin E was heat inactivated after 2 and 60 minutes incubation. All samples were compared to a control containing 100 μM substrate and no enzyme. Following heat inactivation, digestion products were analyzed using a 1100 HPLC system (Agilent) with a C18 column (10 μm bead size, 4.6 mm x 250 mm, Vydac). The HPLC was operated at a flow rate of 1 mL/min with solvent A: 0.1% trifluoroacetic acid in water and solvent B: 0.1% trifluoroacetic acid in 95% acetonitrile. A linear gradient from 5% B to 95% B over 17 minutes was used for peptide separation. Tryptophan fluorescence with excitation at 280 nM and emission at 330 nM was used for detection of the full-length substrate and cleavage products. For quantification of substrate conversion, peak heights were compared between samples and controls.

Peptide cleavage site confirmation by mass spectrometry

Cleavage sites within individual peptides were identified by matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS). 0.5 μL of reaction mixtures were mixed with 0.5 μL of α-Cyano-4-hydroxycinnamic acid (CHCA) matrix and spotted on a MALDI plate. Spectra were then acquired over a mass range of 500-2,500 Daltons using an Axima Performance MALDI-time-of-flight/time-of-flight (TOF/TOF) mass spectrometer (Shimadzu). An average of 200 shots was used for each spectra. All spectra were analyzed using the Shimadzu Biotech Launchpad.

2.6 References

Abd-Elgaliel, W. R., Cruz-Monserrate, Z., Wang, H., Logsdon, C. D., & Tung, C. H. (2013). Pancreatic cancer-associated Cathepsin E as a drug activator. J. Control. Release, 167, 221–227.

Achour, O., Ashraf, Y., Bridiau, N., Kacem, M., Poupard, N., Bordenave-Juchereau, S., Sannier, F., Fruitier-Arnaudin, I., et al. (2016). Alteration of cathepsin D trafficking induced by hypoxia and extracellular acidification in MCF-7 breast cancer cells. Biochimie, 121, 123–130.

Beaujouin, M., Prébois, C., Derocq, D., Laurent-Matha, V., Masson, O., Pattingre, S., Coopman, P., Liaudet-Coopman, E., et al. (2010). Pro-cathepsin D interacts with the extracellular domain of the beta chain of LRP1 and promotes LRP1-dependent fibroblast outgrowth. J. Cell Sci., 123, 3336–3346.

Bergers, G., Javaherian, K., Lo, K., Folkman, J., & Hanahan, D. (1999). Effects of Angiogenesis Inhibitors on Multistage Carcinogenesis in Mice. Science, 284, 808-811.

Bibo-Verdugo, B., O'Donoghue, A. J., Rojo-Arreola, L., Craik, C. S., & García-Carreño, F. (2015). Complementary Proteomic and Biochemical Analysis of Peptidases in Lobster Gastric Juice Uncovers the Functional Role of Individual Enzymes in Food Digestion. Mar. Biotechnol.

Buchholz, M., Braun, M., Heidenblut, A., Kestler, H. a, Klöppel, G., Schmiegel, W., Hahn, S. A., Gress, T. M., et al. (2005). Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions. Oncogene, 24, 6626–6636.

Burster, T., Reich, M., Zaidi, N., Voelter, W., Boehm, B. O., & Kalbacher, H. (2008). Cathepsin E regulates the presentation of tetanus toxin C-fragment in PMA activated primary human B cells. Biochem. Bioph. Res. Co., 377, 1299–1303.

Cappiello, M. G., Wu, Z., Scott, B. B., McGeehan, G. M., & Harrison, R. K. (2004). Purification and characterization of recombinant human cathepsin E expressed in human kidney cell line 293. Protein Expres. Purif., 37, 53–60.

Chain, B. M., Free, P., Medd, P., Swetman, C., Tabor, A. B., & Terrazzini, N. (2005). The Expression and Function of Cathepsin E in Dendritic Cells. J. Immunol., 174, 1791–1800.

Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., & Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. Nature Methods, 6, 786–787.

Costello, E., Greenhalf, W., & Neoptolemos, J. P. (2012). New biomarkers and targets in pancreatic cancer and their application to treatment. Nat. Rev. Gastroentero., 9, 435–444.

Cruz-Monserrate, Z., Abd-Elgaliel, W., & Logsdon, C. (2012). Detection of pancreatic cancer tumours and precursor lesions by cathepsin E activity in mouse models. Gut, 61, 1315–1322.

Cruz-Monserrate, Z., Roland, C. L., Deng, D., Arumugam, T., Moshnikova, A., Andreev, O. a, Reshetnyak, Y. K., Logsdon, C. D. (2014). Targeting pancreatic ductal adenocarcinoma acidic microenvironment. Scientific Reports, 4, 4410.

Dunn, B. M. (2002). Structure and Mechanism of the Pepsin-Like Family of Aspartic Peptidases. Chem. Rev., 102, 4431–4458.

Folkman, J., Watson, K., Ingber, D., & Hanahan, D. (1989). Induction of angiogenesis during the transition from hyperplasia to neoplasia. Nature, 339, 58–61.

Gocheva, V., Wang, H., Gadea, B. B., Shree, T., Hunter, K. E., Garfall, A. L., Berman, T., Joyce, J. A. (2010). IL-4 induces cathepsin protease activity in tumor-associated macrophages to promote cancer growth and invasion. Gene. Dev., 241–255.

Gocheva, V., Zeng, W., Ke, D., Klimstra, D., Reinheckel, T., Peters, C., Hanahan, D., Joyce, J. A. (2006). Distinct roles for cysteine cathepsin genes in multistage tumorigenesis. Gene. Dev., 543–556.

Gopalakrishnan, M. M., Grosch, H.-W., Locatelli-Hoops, S., Werth, N., Smolenová, E., Nettersheim, M., Hasilik, A. (2004). Purified recombinant human prosaposin forms oligomers that bind procathepsin D and affect its autoactivation. Biochem. J., 383, 507–515.

Hezel, A. F., Kimmelman, A. C., Stanger, B. Z., Bardeesy, N., & Depinho, R. a. (2006). Genetics and biology of pancreatic ductal adenocarcinoma. Gene. Dev., 20(10), 1218–1249.

Impens, F., Colaert, N., Helsens, K., Ghesquière, B., Timmerman, E., De Bock, P. J., Chain, B. M. Gevaert, K., et al. (2010). A quantitative proteomics design for systematic identification of protease cleavage events. Mol. Cell. Proteomics , 9, 2327–2333.

Joyce, J. a, Baruch, A., Chehade, K., Meyer-Morse, N., Giraudo, E., Tsai, F.-Y., Greenbaum, D. C., Hanahan, D., et al. (2004). Cathepsin cysteine proteases are effectors of invasive growth and angiogenesis during multistage tumorigenesis. Cancer Cell, 5, 443–453.

Joyce, J. A., & Hanahan, D. (2004). Multiple Roles for Cysteine Cathepsins in Cancer. Cell Cycle, 1516–1519.

Keliher, E. J., Reiner, T., Earley, S., Klubnick, J., Tassa, C., Lee, A. J., Ramaswamy, S., Weissleder, R., et al. (2013). Targeting Cathepsin E in Pancreatic Cancer by a Small Molecule Allows In Vivo Detection. Neoplasia, 15, 684–693.

Kempf, D. J., Sham, H. L., Marsh, K. C., Flentge, C. A., Betebenner, D., Green, B. E., McDonal, E., Norbeck, D. W., et al. (1998). Discovery of Ritonavir, a Potent Inhibitor of HIV Protease with High Oral Bioavailability and Clinical Efficacy. J. Med. Chem., 2623, 602–617.

Kessenbrock, K., Plaks, V., & Werb, Z. (2010). Matrix metalloproteinases: regulators of the tumor microenvironment. Cell, 141, 52–67.

Khalkhali-Ellis, Z., & Hendrix, M. J. C. (2015). Two Faces of Cathepsin D: Physiological Guardian Angel and Pathological Demon. Biology and Medicine, 6.

Laurent-matha, V., Garcia, M., Rochefort, H., Glondu, M., & Coopman, P. (2001). A mutated cathepsin-D devoid of its catalytic activity stimulates the growth of cancer cells. Oncogene, 20, 6920–6929.

Laurent-Matha, V., Lucas, A., Huttler, S., Sandhoff, K., Garcia, M., & Rochefort, H. (2002).

Procathepsin D interacts with prosaposin in cancer cells but its internalization is not mediated by LDL receptor-related protein. Exp. Cell. Res., 277, 210–219.

Laurent-Matha, V., Huesgen, P. F., Masson, O., Derocq, D., Prebois, C., Gary-Bobo, M., Lecaille, F., Liaudet-Coopman, E., et al. (2012). Proteolysis of cystatin C by cathepsin D in the breast cancer microenvironment. FASEB J., 26, 5172–5181.

LeBeau, A. M., Lee, M., Murphy, S. T., Hann, B. C., Warren, R. S., Santos, R. D., Kurhanewicz, J., Craik, C. S., et al. (2013). Imaging a functional tumorigenic biomarker in the transformed epithelium. Proc. Natl. Acad. Sci. USA, 110, 93–98.

Li, H., Li, Y., Cui, L., Wang, B., Cui, W., Li, M., & Cheng, Y. (2014). Monitoring pancreatic carcinogenesis by the molecular imaging of cathepsin E in vivo using confocal laser endomicroscopy. PloS One, 9, 106566.

Lyo, V., Cattaruzza, F., Kim, T. N., Walker, A. W., Paulick, M., Cox, D., Cloyd, J., Kirkwood, K. S. (2012). Active cathepsins B, L, and S in murine and human pancreatitis. Am. J. Physiol.-Gastr. L., 303, 894–903.

Mason, S. D., & Joyce, J. A. (2011). Proteolytic networks in cancer. Trends Cell Bio., 21, 228–237.

Masson, O., Bach, A.-S., Derocq, D., Prébois, C., Laurent-Matha, V., Pattingre, S., & Liaudet-Coopman, E. (2010). Pathophysiological functions of cathepsin D: Targeting its catalytic activity versus its protein binding activity? Biochimie, 92, 1635–1643.

Mohamed, M. M., & Sloane, B. F. (2006). Cysteine cathepsins: multifunctional enzymes in cancer. Nat. Rev. Cancer, 6, 764–775.

Morris, J. P., Wang, S. C., & Hebrok, M. (2010). KRAS, Hedgehog, Wnt and the twisted developmental biology of pancreatic ductal adenocarcinoma. Nat. Rev. Cancer, 10, 683–695.

Mota, F., Kanan, J., Rayment, N., Mould, T., Singer, A., & Chain, B. M. (1997). Cathepsin E Expression by Normal and Premalignant Cervical Epithelium. Am. J. Pathol., 150, 1223–1229.

Nishioku, T., Hashimoto, K., Yamashita, K., Liou, S.-Y., Kagamiishi, Y., Maegawa, H., Katsube, N., Nakanishi, H., et al. (2002). Involvement of cathepsin E in exogenous antigen processing in primary cultured murine microglia. J. Biol. Chem., 277, 4816–4822.

Nolan-stevaux, O., Lau, J., Truitt, M. L., Chu, G. C., Hebrok, M., Hanahan, D., & Ferna, M. E. (2009). GLI1 is regulated through Smoothened- independent mechanisms in neoplastic pancreatic ducts and mediates PDAC cell survival and transformation. Gene. Dev., 1, 24–36.

O'Donoghue, A. J., Eroy-reveles, A. A., Knudsen, G. M., Ingram, J., Zhou, M., Statnekov, J. B., Greninger, A. L., Craik, C. S., et al. (2012). Global identification of peptidase specificity by multiplex substrate profiling. Nature Methods, 9, 1095-1100.

Ostermann, N., Gerhartz, B., Worpenberg, S., Trappe, J., & Eder, J. (2004). Crystal structure of an activation intermediate of cathepsin E. J. Mol. Biol., 342, 889–899.

Rao-Naik, C., Guruprasad, K., Batley, B., Rapundalo, S., Hill, J., Blundell, T., Kay, J., Dunn, B. M. (1995). Exploring the Binding Preferences/Specificity in the Active Site of Human Cathepsin E. Proteins, 181, 168–181.

Reid, A., Valler, M., & Kay, J. (1986). Immunolocalisation of cathepsin D in normal and neoplastic human tissues. J. Clin. Pathol., 1323–1330.

Richter, C., Tanaka, T., & Yada, R. Y. (1998). Mechanism of activation of the gastric aspartic proteinases : pepsinogen, progastricsin and prochymosin. Biochem. J., 490, 481–490.

Ringel, J., Jesnowski, R., Moniaux, N., Lüttges, J., Ringel, J., Choudhury, A., Batra, S. K., Löhr, M., et al. (2006). Aberrant expression of a disintegrin and metalloproteinase 17/tumor necrosis factor-alpha converting enzyme increases the malignant potential in human pancreatic ductal adenocarcinoma. Cancer Res., 66, 9045–9053.

Sakai, H., Saku, T., Yuzo, K., & Yamamolo, K. (1989). Quantitation and immunohistochemical localization of cathepsins E and D in rat tissues and blood cells. Biochim. Biophys. Acta, 991, 367–375.

Sastradipura, D. F., Nakanishi, H., Tsukuba, T., Nishishita, K., Sakai, H., Kato, Y., Gotow, T., Yamamoto, K. (2002). Identification of Cellular Compartments Involved in Processing of Cathepsin E in Primary Cultures of Rat Microglia. J. Neurochem., 70, 2045–2056.

Sevenich, L., Bowman, R. L., Mason, S. D., Quail, D. F., Rapaport, F., Elie, B. T., Joyce, J. A., et al. (2014). Analysis of tumour- and stroma-supplied proteolytic networks reveals a brain-metastasis-promoting role for cathepsin S. Nat. Cell Biol., 16, 876–888.

Sevenich, L., & Joyce, J. A. (2014). Pericellular proteolysis in cancer. Gene. Dev., 2331–2347.

Tsukuba, T., Okamoto, K., Okamoto, Y., Yanagawa, M., Kohmura, K., Yasuda, Y., Uchi, H., Nakayama, K. I., et al. (2003). Association of Cathepsin E Deficiency with Development of Atopic Dermatitis. J. Biochem., 134, 893–902.

Tsukuba, T., Yamamoto, S., Yanagawa, M., Okamoto, K., Okamoto, Y., Nakayama, K. I., Kadowaki, T., Yamamoto, K. (2006). Cathepsin E-deficient mice show increased susceptibility to bacterial infection associated with the decreased expression of multiple cell surface Toll-like receptors. J. Biochem., 140, 57–66.

Ullmann, R., Morbini, P., Halbwedl, I., Bongiovanni, M., Gogg-Kammerer, M., Papotti, M., Gabor, S., Popper, H. H., et al. (2004). Protein expression profiles in adenocarcinomas and squamous cell carcinomas of the lung generated using tissue microarrays. J. Pathol., 203, 798–807.

Uno, K., Azuma, T., Nakajima, M., Yasuda, K., Hayakumo, T., Mukai, H., Sakai, T., Kawai, K. (2000). Clinical significance of cathepsin E in pancreatic juice in the diagnosis of pancreatic ductal adenocarcinoma. J. Gastroen. Hepatol., 15, 1333–1338.

Yasuda, Y., Kageyama, T., Akamine, a., Shibata, M., Kominami, E., Uchiyama, Y., & Yamamoto, K. (1999). Characterization of New Fluorogenic Substrates for the Rapid and Sensitive Assay of Cathepsin E and Cathepsin D. J. Biochem., 125, 1137–1143.

Yasuda, Y., Kohmura, K., Kadowaki, T., Tsukuba, T., & Yamamoto, K. (2005). A new selective substrate for cathepsin E based on the cleavage site sequence of alpha2-macroglobulin. Biol. Chem., 386, 299–305.

Zaidi, N., Herrmann, T., Voelter, W., & Kalbacher, H. (2007). Recombinant cathepsin E has no proteolytic activity at neutral pH. Biochem. Biophys. Res. Co., 360, 51–55.

Zaidi, N., & Kalbacher, H. (2008). Cathepsin E: a mini review. Biochem. Biophys. Res. Co., 367, 517–522.

A

kDa

72 —

55 —

43 —

Brefeldin A   -   +

B

Lysate Media Lysate Media Lysate Media

kDa

72 —

55 —

43 —

pH 7.4   pH 3.5   PNGaseF

C

CTSE   EEA1   Merge
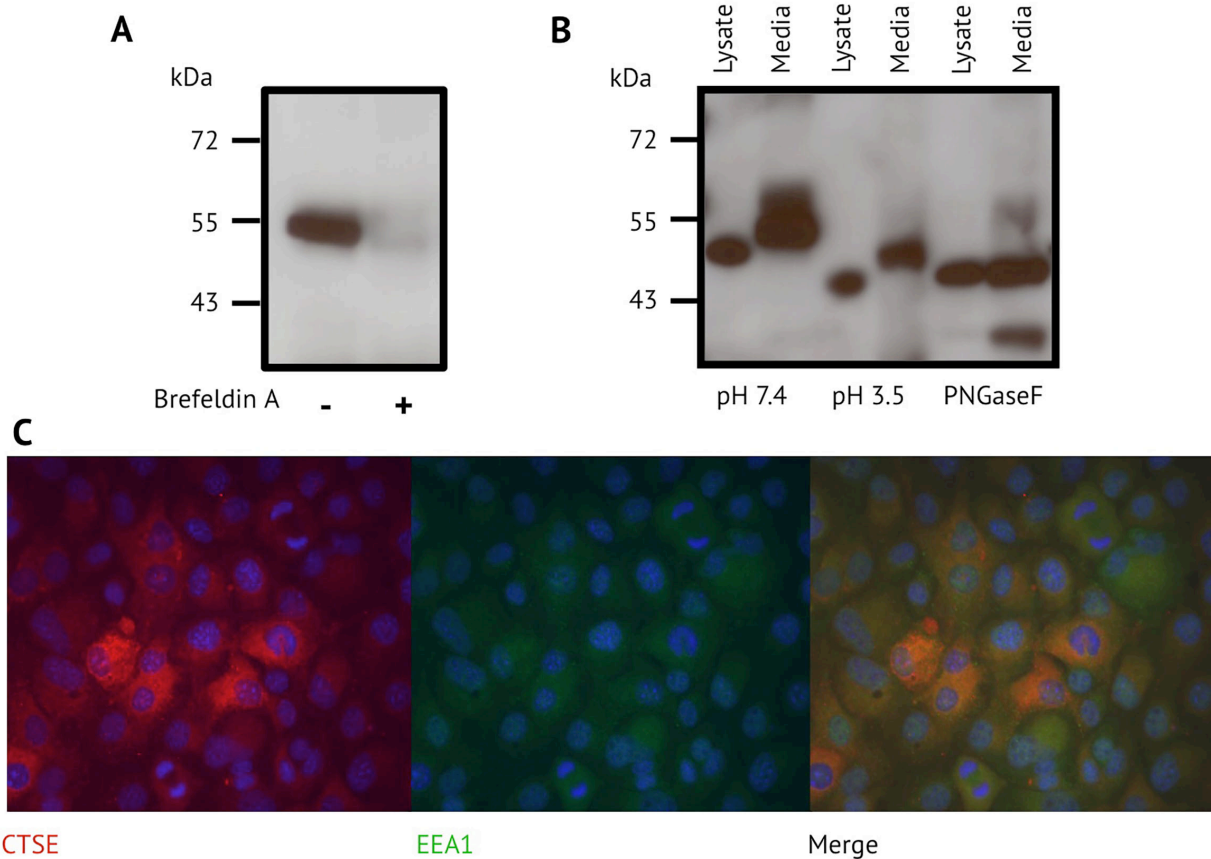
Figure 2.1 Expression and localization of cathepsin E in a mouse PDAC cell line. (A) Western blot analysis of cathepsin E in conditioned media with and without brefeldin A treatment. Size markers correspond to kilodaltons (kDa). (B) Western blot analysis of cathepsin E in the cell lysate and conditioned media treated with acid and the deglycosylase PNGaseF. (C) Immunofluorescence of cathepsin E and the endosomal marker EEA1. DAPI is shown in blue.
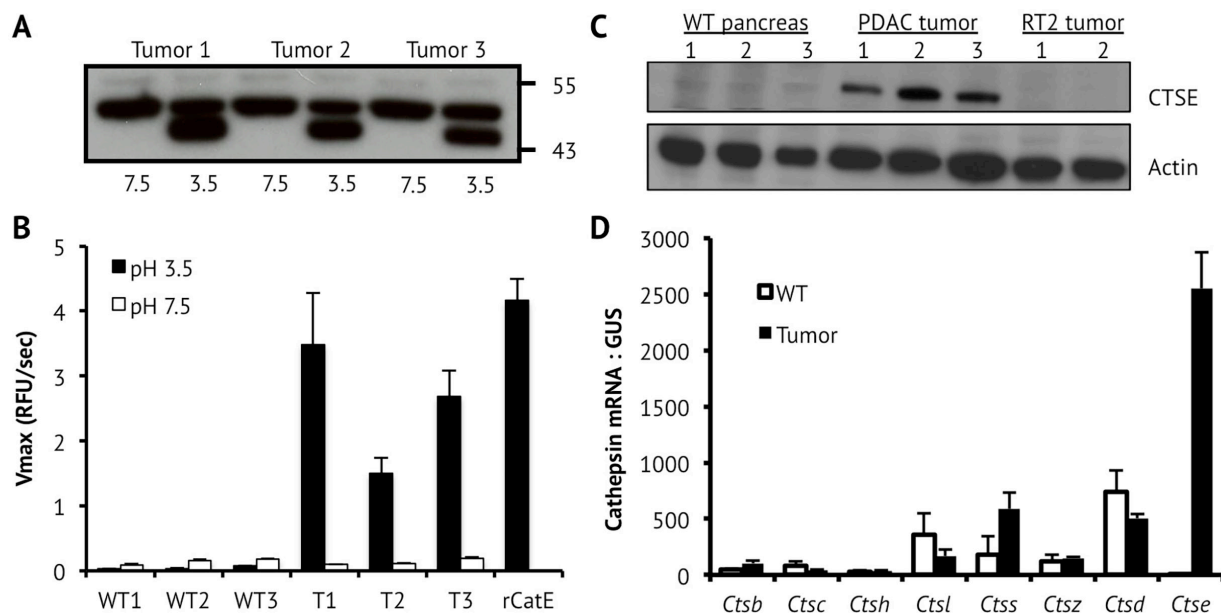
Figure 2.2 Cathepsin E expression and activity in mouse PDAC tumors. (A) Western blot analysis of cathepsin E from three tumors before and after acid treatment. (B) Cathepsin E activity in tumors and pancreatic tissue from wild type (WT) littermates. Cleavage velocity is given in relative fluorescence units per second (RFU/sec). (C) Immunoblot analysis of cathepsin E expression in pancreatic tissue from WT, *p48-Cre; Kras^{G12D}; Trp53^{f/f}*, and RIP1-Tag2 mice. (D) Quantitative PCR analysis of cathepsin mRNA levels in pancreatic tissue from WT and PDAC animals. GUS mRNA levels were used for normalization.
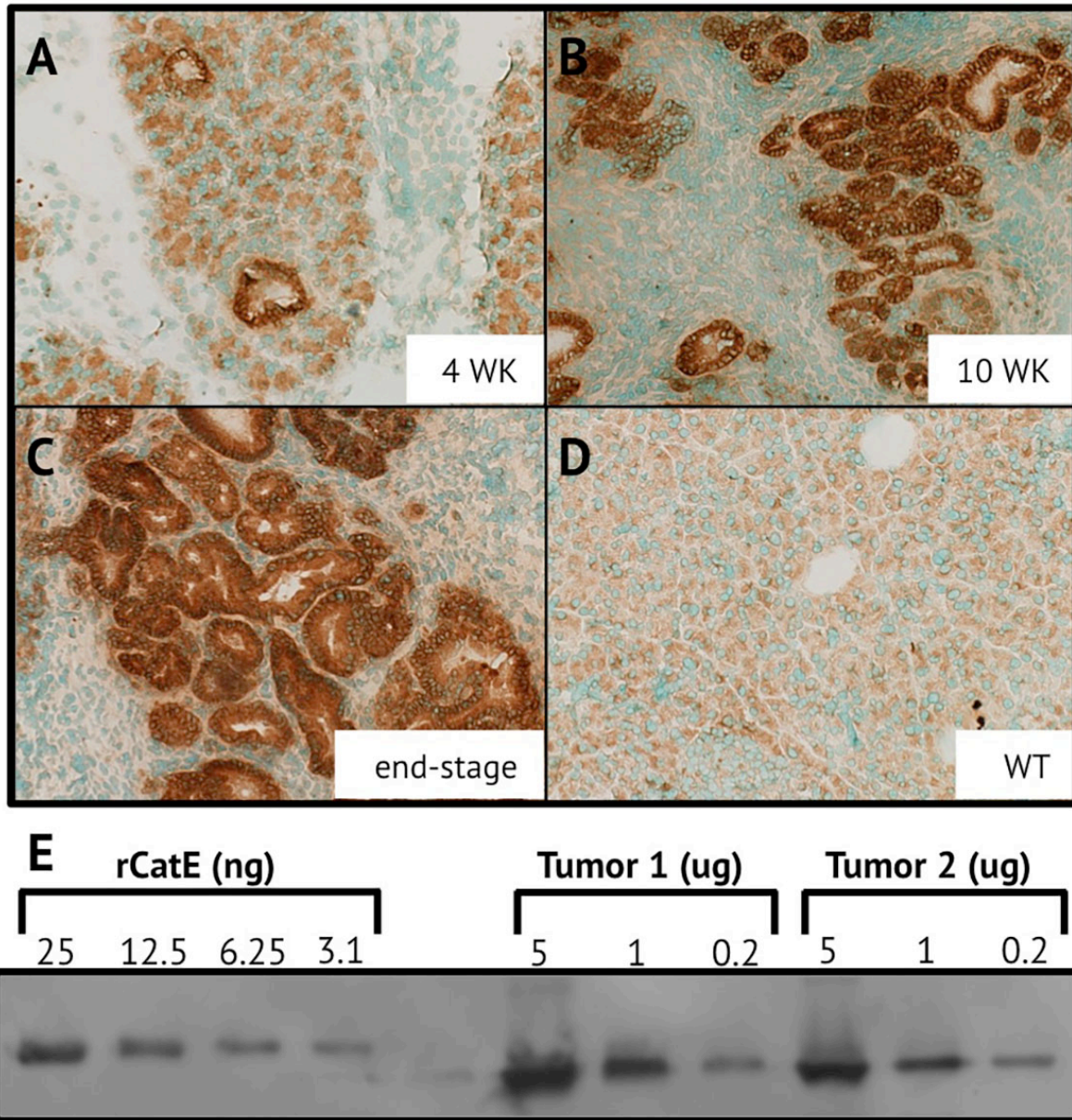
Figure 2.3 Immunohistochemical staining of cathepsin E in pancreatic tissue from *p48-Cre; Kras^{G12D}; Trp53$^{f/f}$* mice after (A) 4 weeks, (B) 10 weeks, (C) and in end-stage (D) and WT mice. (E) Semi-quantitative western blot analysis of cathepsin E from tumor lysate of end-stage mice.
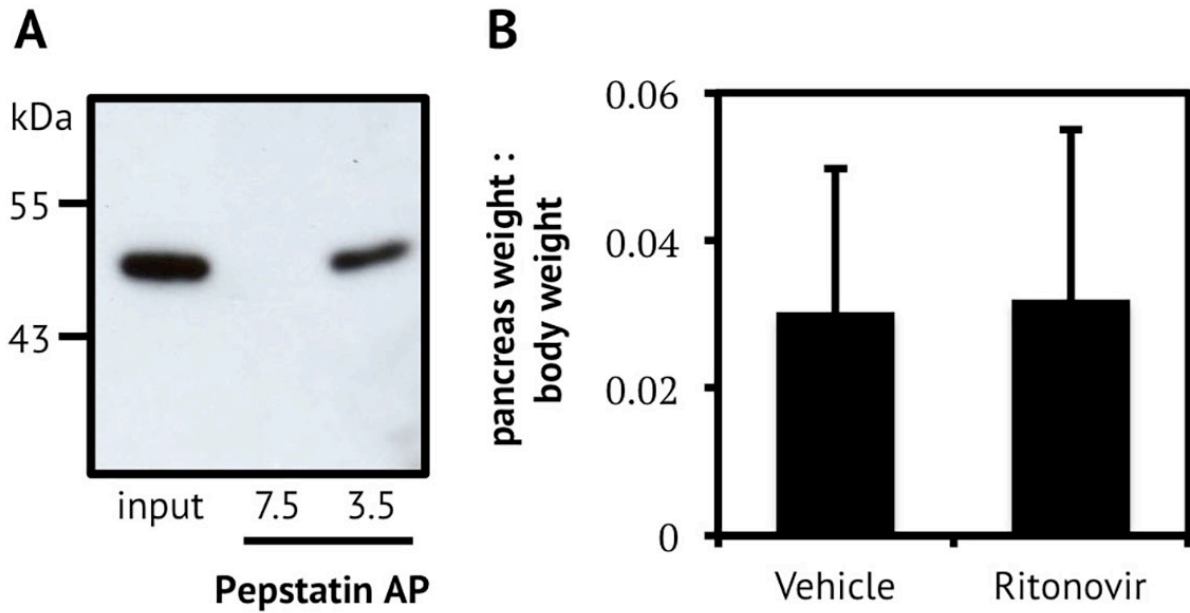
Figure 2.4 Cathepsin E inhibitor binding in tumor lysate and inhibition in PDAC mice. (A) Affinity purification (AP) of cathepsin E from PDAC tumors using pepstatin-agarose beads followed by western blot analysis. Acid treatment was required for affinity purification of cathepsin E. (B) Effect of ritonavir treatment on tumor burden in PDAC mice.
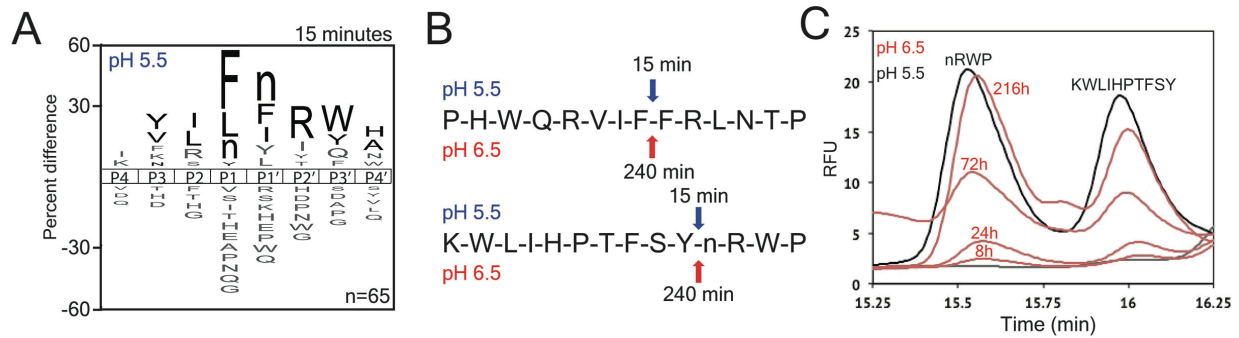
Figure 2.5 Analysis of procathepsin E activity. (A) iceLogo generated from the cathepsin E cleavage events detected after 15 minutes using the MSP-MS assay. To profile mature cathepsin E the MSP-MS assay was performed at pH 5.5. In the iceLogo n corresponds to norleucine. (B) The cleavage sites within two peptides by mature cathepsin E (pH 5.5 – blue) and procathepsin E (pH 6.5 – red). The time at which cleavage was first observed is shown above the arrows indicating the site of cathepsin E cleavage. (C) HPLC chromatograms of peptide cleavage products, which have amino acid sequences shown above each peak, following incubation with procathepsin E (pH 6.5 – red) and mature cathepsin E (pH 5.5 – black). Abundance of cleavage products was monitored using tryptophan fluorescence with excitation at 280 nm and emission at 330 nm.
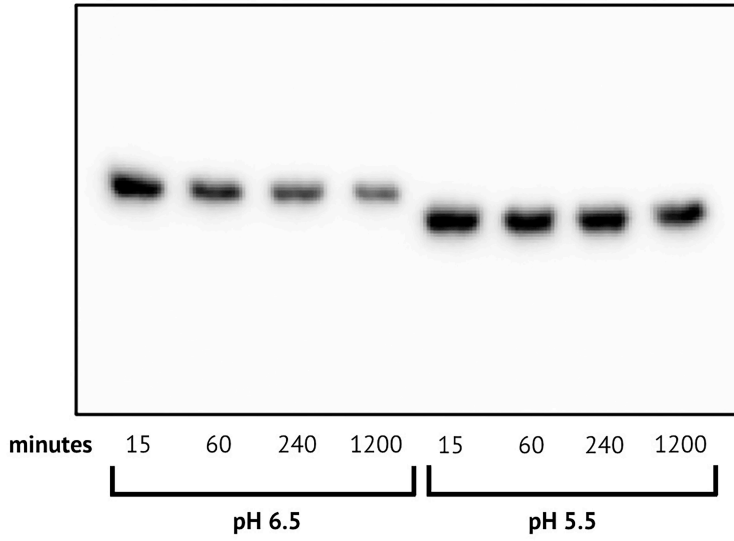
Supplementary Figure 2.1 Cathepsin E activation. Western blot analysis of recombinant cathepsin E incubated for up to 1200 minutes in pH 5.5 or pH 6.5 acetate buffer.
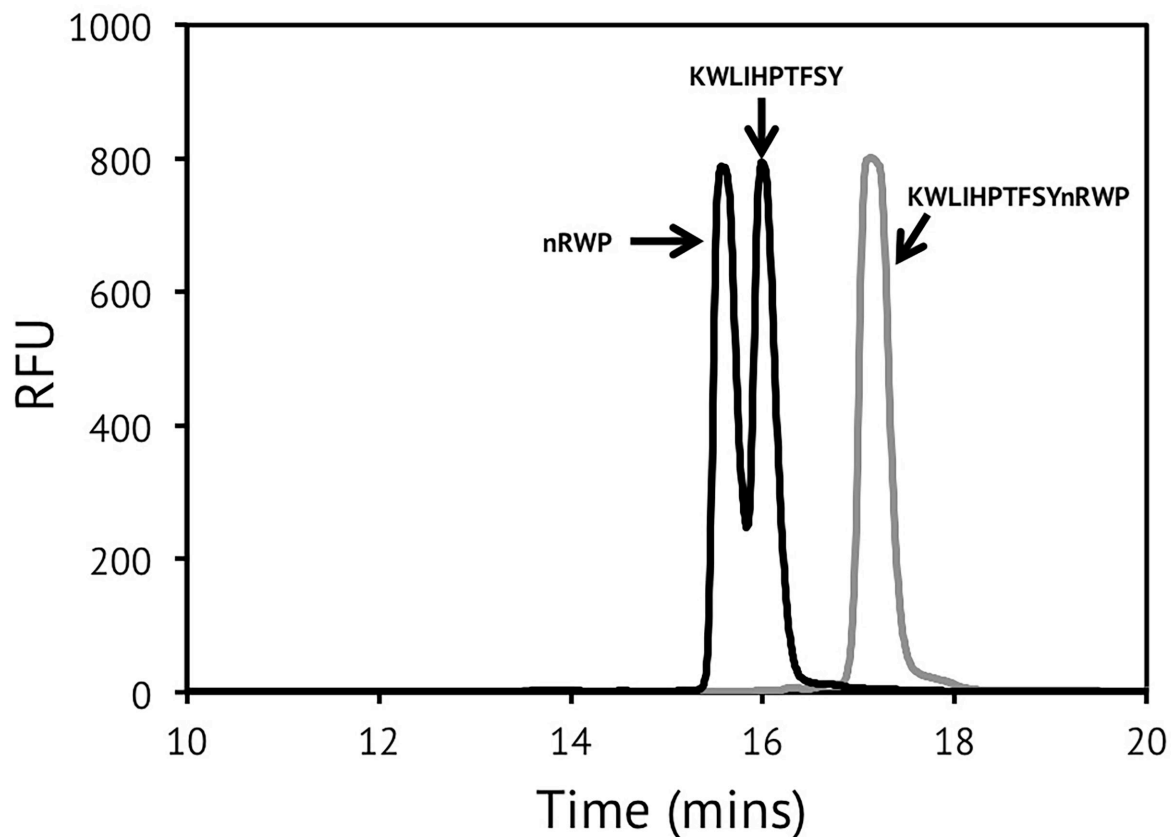
Supplementary Figure 2.2 Peptide cleavage by mature cathepsin E. Peptide HPLC chromatogram when incubated for one hour in pH 5.5 acetate buffer (gray) or with 2 nM cathepsin E in pH 5.5 acetate buffer (black). Cleavage products and parent peptide abundance were quantified using tryptophan fluorescence with excitation at 280 nm and emission at 330 nm.
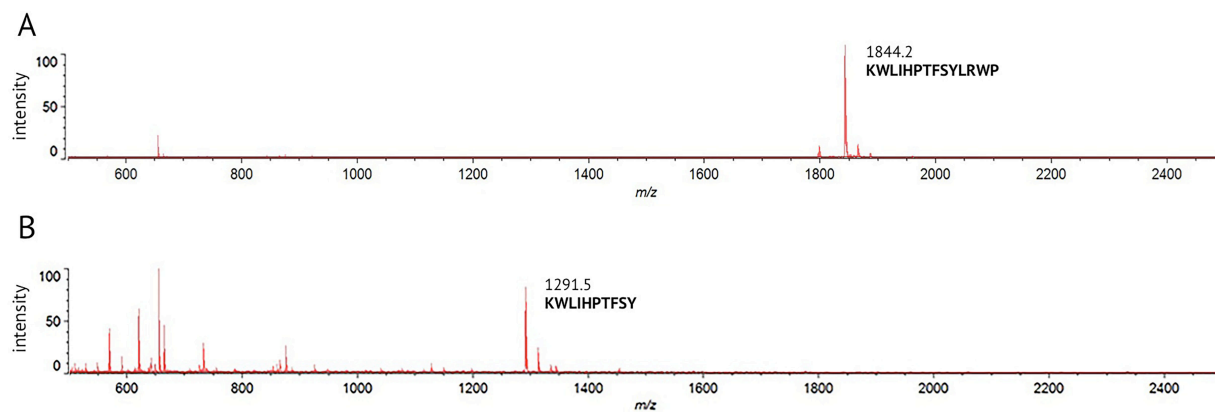
Supplementary Figure 2.3 Identification of peptide cleavage site by MALDI-MS. (A) One hour incubation in pH 5.5 acetate buffer. (B) One hour incubation in pH 5.5 acetate buffer with 2 nM cathepsin E.

Chapter 3. Global protease activity profiling provides differential diagnosis of pancreatic cysts

3.1 Translational relevance: With advances in abdominal imaging technologies, the incidental detection of pancreatic cysts continues to rise. However, there remains a lack of accurate molecular diagnostics for differentiating benign cystic lesions from those that can progress into pancreatic cancer. This has led to a dramatic increase in the number of potentially unnecessary pancreatic resections, which are associated with high rates of morbidity. Using a global and unbiased protease-activity profiling approach and patient cyst fluid, we determined that the activities of the aspartyl proteases gastricsin and cathepsin E accurately differentiate premalignant mucinous cysts from benign nonmucinous cysts. In particular, analysis of gastricsin activity demonstrated 93% sensitivity and 100% specificity for differentiating mucinous lesions. Our simple and direct fluorescence-based approach for stratification of pancreatic cysts significantly outperformed the most widely used molecular biomarker – carcinoembryonic antigen (CEA) – and can be readily translated into an actionable diagnostic assay to help improve clinical management of these challenging lesions.

3.2 Abstract

Purpose: Pancreatic cysts are estimated to be present 2-3% of the adult population. Unfortunately, current diagnostics do not accurately distinguish benign cysts from those that can progress into invasive cancer. Missregulated pericellular proteolysis is a hallmark of malignancy, and therefore, we used a global approach to discover protease activities that differentiate benign nonmucinous cysts from premalignant mucinous cysts.

Experimental Design: We employed an unbiased and global protease profiling approach to discover protease activities in 23 cyst fluid samples. The distinguishing activities of select proteases was confirmed in 110 samples using specific fluorogenic substrates and required less than 5 µL of cyst fluid.

Results: We determined that the activities of the aspartyl proteases gastricsin and cathepsin E are highly increased in fluid from mucinous cysts. Immunohistochemical analysis revealed that gastricsin expression was associated with regions of low-grade dysplasia, whereas cathepsin E expression was independent of dysplasia grade. Gastricsin activity differentiated mucinous from nonmucinous cysts with a specificity of 100% and a sensitivity of 93%, whereas cathepsin E activity was 92% specific and 70% sensitive. Gastricsin significantly outperformed the most widely used molecular biomarker, carcinoembryonic antigen (CEA), which demonstrated 94% specificity and 65% sensitivity. Combined analysis of gastricsin and CEA resulted in a near perfect classifier with 100% specificity and 98% sensitivity.

Conclusions: Quantitation of gastricsin and cathepsin E activities accurately distinguished mucinous from nonmucinous pancreatic cysts and has the potential to replace current diagnostics for analysis of these highly prevalent lesions.

3.3 Introduction

The detection of pancreatic cysts has increased dramatically due to the rising use of high-resolution abdominal imaging. Pancreatic cysts are incidentally detected in 13-45% of patients evaluated by magnetic resonance imaging and 2% of patients evaluated by computed tomography (1–3). The most frequently detected pancreatic cysts include intraductal papillary mucinous neoplasms (IPMNs), mucinous cystic neoplasms (MCNs), pseudocysts, and serous

cystadenomas (SCAs) (4). Both IPMNs and MCNs, which are collectively referred to as mucinous cysts, may develop foci of high-grade dysplasia or cancer (5). At the time of resection, ~30% of IPMNs and ~15% of MCNs contain invasive cancer (6,7). Pseudocysts and SCAs, which are both nonmucinous, rarely undergo malignant degeneration and are considered benign lesions that typically do not require resection or continued surveillance. Clinical decision making for pancreatic cysts relies largely on radiographic and clinical features, augmented by analysis of cyst fluid collected by endoscopic ultrasound with fine needle need aspiration (EUS-FNA) (8). Unfortunately, with current clinical guidelines, distinguishing nonmucinous from mucinous cysts remains a challenge. The preoperative diagnosis of mucinous cysts is incorrect in up to 30% of cases and benign cysts are often resected, exposing patients to an unnecessary risk for morbidity (9–12).

As abdominal imaging remains unable to accurately differentiate pancreatic cyst types, there has been considerable effort towards developing improved diagnostic biomarkers. Most of these biomarkers utilize cyst fluid collected by EUS-FNA. CEA is the most widely investigated biomarker and is 60-80% accurate for differentiating mucinous from nonmucinous cysts (13,14). *KRAS* mutations occur in more than 90% of pancreatic cancers and are frequently observed in mucinous cysts (15,16). Analysis of cyst fluid DNA revealed that *KRAS* mutations are 100% specific, but only 50% sensitive for diagnosing a mucinous cyst (17). Similarly, analysis of mutations in the oncogene *GNAS* are specific for diagnosing IPMNs, but suffer from low sensitivity (18). A variety of other cyst fluid biomarkers have also been explored (19–23); however, CEA remains the only widely applied molecular biomarker for differentiating mucinous from nonmucinous cysts.

Proteases mediate a variety of critical processes in cancer, including invasion of the basement membrane via cleavage of extracellular matrix proteins and promotion of oncogenic signaling pathways through activation of growth factors and receptor tyrosine kinases (24,25). In pancreatic cancer, members of the cathepsin family of endolysosomal proteases are upregulated and found extracellularly. Aberrant secretion leads to cleavage of extracellular substrates, driving increased cellular invasion (26). Either genetic deletion or pharmacological inhibition of cysteine cathepsin activity decreases tumor progression and invasion (27,28).

Gene expression profiling studies of IPMNs and MCNs indicate overexpression of a range of proteases (29–31). Furthermore, analysis of protein expression in cyst fluid showed substantial differences in the abundance of pancreatic proteases and their cognate inhibitors between cyst types. The serine protease inhibitor SPINK1 was recently investigated as a biomarker for differentiating benign from malignant cysts (32,33). Collectively, these results suggest that there may be altered levels of proteolytic activity between mucinous and nonmucinous cysts and that these differences could be exploited to distinguish the type of lesion and its associated malignant potential.

In the current study, we applied a global protease profiling technology to discover proteolytic activity markers for differentiating mucinous from nonmucinous cysts. Using this approach, we identified enhanced aspartyl protease activity in mucinous cysts, due to upregulation of gastricsin and cathepsin E. We characterized the localization of both aspartyl proteases within the dysplastic tissue surrounding the mucinous cysts and determined that gastricsin expression was dependent on the degree of dysplasia. Lastly, highly selective fluorescent substrates for gastricsin and cathepsin E both confirmed their upregulated activities and outperformed CEA for differentiating mucinous from nonmucinous pancreatic cystic lesions.

3.4 Materials and methods

Patients and sample acquisition

Pancreatic cyst fluid samples were collected from preconsented patients under institutional review board approved protocols and in accordance with U. S. Common Rule at the University of California San Francisco, the University of Pittsburgh Medical Center, Indiana University School of Medicine, and Stanford University School of Medicine. Patient information is summarized in Table S1. All patients included in our study underwent surgical resection of their cystic lesion and have a pathologically confirmed diagnosis. The highest grade of dysplasia observed during pathological evaluation of each cystic lesion is reported. Samples were collected either at the time of surgical resection or during diagnostic endoscopic ultrasound prior to resection of the cystic lesion. Cyst fluid samples were split into 100 μL aliquots and frozen to -80 °C within 60 minutes of collection. Samples underwent at most two freeze-thaw cycles prior to experimental analysis. Total cyst fluid protein concentration was determined by the bicinchoninic acid assay. CEA levels were evaluated for the majority of samples, but were unavailable in 21 cases due to limited cyst fluid volume.

Multiplex substrate profiling by mass spectrometry (MSP-MS) assay

The MSP-MS assay was performed as previously described (34). Cyst fluid was diluted to 100 μg/mL in assay buffer (either pH 7.5 phosphate buffer or pH 3.5 acetate buffer) and pre-incubated for 10 minutes. For analysis of protease inhibitor sensitivity, 1 mM AEBSF (Sigma, A8456), 2 μM E-64 (Sigma, E3132), 2 μM pepstatin (Sigma, P5318), 2 mM 1,10-phenanthroline (Sigma, 131337), or DMSO were included in pre-incubation. The 228 tetradecapeptide library

was split into two pools and diluted in assay buffer to a concentration of 1 μM of each peptide. 75 μL of diluted cyst fluid and peptide pools were then combined and incubated at room temperature. 30 μL aliquots were removed after 15 and 60 minutes, protease activity quenched with 8 M guanidinium hydrochloride, and flash-frozen in liquid N_2. For recombinant gastricsin (R&D Systems, 6186-AS), cathepsin D (R&D Systems, 1014-AS), and cathepsin E (R&D Systems, 1294-AS), the MSP-MS assay was performed as described above with slight modifications: 10 nM of recombinant protease in pH 3.5 acetate buffer was used and aliquots were removed after 15, 60, and 240 minutes.

Prior to peptide cleavage site identification by mass spectrometry, samples were desalted using C18 tips (Rainin). Mass spectrometry analysis was carried out with an LTQ Orbitrap XL mass spectrometer coupled to a 10,000 psi nanoACQUITY Ultra Performance Liquid Chromatography (UPLC) System (Waters) for peptide separation by reverse phase liquid chromatography (RPLC). Peptides were separated over a C18 column (Thermo) and eluted by applying a flow rate of 300 nL/min with a 65-minute linear gradient from 2-30% acetonitrile. Survey scans were recorded over a 325-1500 m/z range and the six most intense precursor ions were fragmented by collision-induced dissociation (CID) for MS/MS.

Raw mass spectrometry data was processed to generate peak lists using MSConvert. Peak lists were then searched in Protein Prospector v. 5.10.0 (35) against a custom database containing the sequences from the 228 tetradecapeptide library. Searches used a mass accuracy tolerance of 20 ppm for precursor ions and 0.8 Da for fragment ions. Variable modifications included N-terminal pyroglutamate conversion from glutamine or glutamate and oxidation of tryptophan, proline, and tyrosine. Searches were subsequently processed using the MSP-xtractor software (http://www.craiklab.ucsf.edu/extractor.html), which extracts the peptide cleavage site and

spectral counts of the corresponding cleavage products. Spectral counts were used for the relative quantification of peptide cleavage products.

Proteomic analysis of cyst fluid samples

Cyst fluid samples were processed for proteomic analysis using a standard protocol. Briefly, 8 μg of cyst fluid protein was denatured in 40 μL of 6 M urea. Disulfide bonds were reduced with 10 mM dithiothreitol and free thiols were subsequently alkylated with 12.5 mM iodoacetamide. Samples were then diluted to with 25 mM ammonium bicarbonate to 2 M urea and digested with 100 ng trypsin for 16 hours at 37 °C. Following trypsin digestion, samples were desalted with C18 tips (Rainin), dried, and resuspended in 0.1% formic acid. Triplicate LC-MS/MS analysis of all samples was performed as described above and details for this and protein identification are provided in the Supplemental Methods and Materials.

Label-free quantitation was used to compare relative abundance of the three aspartyl proteases identified in cyst fluid samples. The Skyline software package was used to obtain extracted ion chromatograms and peak areas for precursor ions from the aspartyl proteases (36). To correct for potential differences in protein loading between runs, peak areas were normalized by the median peak area of all fragmented ions from that run. The average peak area of the precursor ions from a given aspartyl protease was then used to estimate the abundance in each cyst fluid sample.

Western blots of gastricsin and cathepsin E

Cyst fluid protein (2 μg) or recombinant protease (20 ng) was pre-incubated for 30 minutes in either pH 7.5 phosphate buffer or pH 3.5 acetate buffer. Samples were then subjected to

electrophoresis on a 10% NuPAGE Bis-Tris gel. Proteins were transferred to polyvinylidene fluoride membranes and blocked in Tris-buffered saline with 0.1% Tween (TBS-T) and 5% (w/v) non-fat dry milk for 2 hours at room temperature. Membranes were then incubated with either rabbit anti-gastricsin antibody (1:500; Abcam, ab104595) or rabbit anti-cathepsin E antibody (1:1,000; Abcam, ab49800) for 1 hour at room temperature. Following a wash in TBS-T, horseradish peroxidase (HRP)-conjugated secondary antibody (1:15,000; Abcam, ab97051) was applied for 2 hours at room temperatures. Proteins were detected with the enhanced chemiluminescence (ECL) detection system (Thermo).

Animal strains

The following mice strains were used: *Ptf1a-Cre* (gift of Christopher Wright, Vanderbilt University, Nashville, Tennessee, USA), *LSL-Kras^{G12D}* (gift of Dave Tuveson, Cold Spring Harbor Laboratory, USA), *Brg1$^{f/f}$* (gift of David Reisman, University of Florida, USA with permission of Pierre Chambon). Mice were crossed on a mixed background. The UCSF Institutional Care and Use of Animals Committee (IACUC) approved all mouse experiments.

Immunohistochemical analysis of pancreatic tissue

Tissue samples were obtained from patients who underwent resection of pancreatic cystic lesions at UCSF. Gastricsin and cathepsin E immunohistochemistry assays were developed and performed on a Ventana Discovery Ultra automated slide stainer (Ventana Medical Systems). In brief, formalin-fixed, paraffin-embedded (FFPE) samples (4 μm sections) were deparaffinized using EZPrep (Ventana Medical Systems) followed by treatment with antigen retrieval buffer (Ventana Medical Systems, 950-124). Specimens were incubated with either goat anti-gastricsin

antibody (1:300; Santa Cruz, sc-51185) or goat anti-cathepsin E antibody (1:200; Santa Cruz, sc-6508) for 32 minutes at 36 °C. OmniMap anti-goat secondary antibody (Ventana Medical Systems, 760-4647) was then applied for 16 minutes before employing a DAB detection kit (Ventana Medical Systems, 760-500). All samples were counterstained with haematoxylin and Bluing Reagent (Ventana Medical Systems, 760-2037). H&E staining of tissue sections was performed using standard protocols.

Mouse pancreatic tissue samples were collected from 8 *Ptf1a-Cre; LSL-Kras*^{G12D}*; Brg1*$^{f/f}$ animals between 3 and 40 weeks of age. FFPE samples (5 μm sections) were deparaffinized with xylene and subsequently rehydrated. Sections were either subjected to H&E staining or heat-induced epitope retrieval with Citra buffer (BioGenex; HK086). Primary antibodies (goat anti-mouse) for cathepsin E (1:1,000; R&D Systems; AF1130) and gastricsin (1:1,000; Santa Cruz; sc-51188) were incubated with sections overnight at 4 °C. Anti-goat secondary antibody (1:200; Vector Labs; BA-9500) was then added to sections for 1 hour at room temperature. ABC (Vector Labs; PK-6100) and DAB kits (Vector Labs; SK-4100) were employed for detection. Sections were counterstained with haematoxylin and incubated in 0.25% ammonium hydroxide for bluing.

Peptide synthesis

Synthesis of internally quenched fluorescent peptides was conducted using standard Fmoc solid-phase peptide synthesis on a Syro II automated synthesizer (Biotage). Details are included in the Supplemental Methods and Materials.

Fluorescent protease activity assays

All fluorescent protease activity assays were performed in triplicate in black, round-bottom 384 well plates. Assays were run for 1 hour in 15 µL of acetate buffer with 0.01% Tween. The pH of the acetate buffer was adjusted to promote activity of either aspartyl protease (pH 3.5 for the cathepsin E substrate and pH 2.0 for the gastricsin substrate). 10 µM of substrate was used for all assays (unless otherwise indicated) and was incubated with either 10 nM of recombinant protease or 50 µg/mL of cyst fluid protein. For kinetic analysis of gastricsin activity, the substrate concentration ranged from 0.1-25 µM. Fluorescent substrate cleavage was monitored with a Biotek Synergy HT plate reader using excitation and emission wavelengths of 328 nm and 393 nm, respectively. Selectivity of the recombinant proteases was assessed by comparing the initial velocity of substrate hydrolysis in relative fluorescent units per second (RFU/sec). For cyst fluid samples, we compared the change in endpoint RFU relative to wells that contained substrate, but no cyst fluid.

Statistical analysis and data presentation

A two-tailed Mann-Whitney U test was used to compare the differences in CEA abundance, gastricsin activity, and cathepsin E activity between mucinous and nonmucinous cysts. Univariate and multivariate logistic regression models were used for cyst prediction. Receiver operating characteristic (ROC) curves and Youden's J statistic were employed to identify the optimal cutoff. All mass spectrometry data (spectral counts and peak areas) was log2 transformed and analyzed with unpaired two-tailed t-tests. GraphPad Prism was used to fit kinetic data and generate scatter plots and bar charts. Volcano plots, heat maps, venn diagrams, ROC curves, and logistic regression models were generated in RStudio v. 0.98.1091. iceLogo

software was used to visualize patterns in peptide cleavage sites at ±4 positions away from the scissile bond (37).

Supplemental Methods and Materials

Peptide synthesis

Peptide synthesis reactors were charged with Wang resin preloaded with Fmoc-Lys(dinitrophenol) (Anaspec, A23856). Coupling reactions were conducted on a 0.0125 mmol scale in 500 µL of DMF with 5 equivalents of Fmoc amino acid (Anaspec), 4.9 equivalents of HCTU (Anaspec), and 20 equivalents of N-methylmorpholine (NMM). Reactions were run for 8 minutes while shaking at ambient temperature and atmosphere. Fmoc deprotection was conducted by incubation with 500 µL 40% 4-methlypiperidine in DMF for 3 minutes, followed by 500 µL 20% 4-methlypiperidine in DMF for 10 minutes. Substrates were capped with a C-terminal Lys-(7-methoxycoumarin-4-acetic acid)-OH (EMD Millipore, 852095) followed by two D-Arg residues to enhance solubility. Peptides were removed from resin by incubating with 500 µL of cleavage solution (95% trifluoroacetic acid, 2.5% water, 2.5% Triisopropylsilane) and shaking for 1h. Crudes were precipitated in 10 mL cold diethyl ether and isolated by centrifugation. Peptides were purified by HPLC and the correct mass was confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

LC-MS/MS analysis of cyst fluid samples and protein identification

Triplicate LC-MS/MS analysis of all samples was performed on an LTQ Orbitrap XL mass spectrometer and UPLC system described previously. Peptides were separated by RPLC over a C18 column (Thermo). A 60-minute linear gradient from 2-15% acetonitrile with a flow rate of

300 nL/min was used for elution of peptides. Survey scans were recorded over a 325-1500 m/z range. The six most intense precursor ions from each survey scan were fragmented by collision-induced dissociation (CID) for MS/MS analysis.

Peak lists were generated using an in-house software called PAVA and searched in Protein Prospector v. 5.10.0 (35). Peak lists were searched against all human protein sequences in the SwissProt database (downloaded December 1, 2015 with 549,832 sequence entries). This database was concatenated with a fully randomized set of entries to estimate the false discovery rate (FDR). For database searches, peptides sequences were matched to tryptic peptides with up to two missed cleavages. Carbamidomethylation of cysteine residues was used as a constant modification and variable modifications included oxidation of methionine, N-terminal pyroglutamate from glutamine, N-terminal acetylation, and loss of N-terminal methionine. The mass accuracy tolerance was set to 20 ppm for precursor ions and 0.8 Da for fragment ions. Protein Prospector score thresholds of 22 for the protein score and 15 for the peptide score were used and yielded a FDR of less than 1%. All reported proteins were identified by at least 2 unique peptides.

3.5 Results

Global protease activity profiling of patient cyst fluid

To identify differences in proteolytic activity between mucinous and nonmucinous cysts we used our MSP-MS assay, which is a global and unbiased substrate-based protease profiling approach (34). In the MSP-MS assay, a physicochemically diverse library of 228 tetradecapeptide substrates is incubated with a protease-containing sample of interest and tandem mass spectrometry is used to monitor protease-derived peptide cleavage products. We have previously

validated this assay through analysis of all classes of protease and used it to develop selective substrate probes (38–40).

Using the MSP-MS assay, we profiled 16 mucinous and 7 nonmucinous cyst fluid samples. To capture a broad range of protease activities, we performed the assay under acidic conditions and at neutral pH. At pH 7.5, we detected a total of 1117 unique peptide cleavages among the patient sample set (Fig. 1A). Only 7 peptide cleavages met our selectivity criteria for differentiating mucinous from nonmucinous cysts (+/- 1 $\log_2$(mucinous/nonmucinous), $p < 0.05$). 6 of these cleavages were enriched in nonmucinous cysts, and overall, there was a slight trend toward increased proteolytic activity in fluid from nonmucinous cysts (Fig. S1A). When the same samples were assayed at pH 3.5, a total of 691 peptide cleavages were detected, and we observed increased proteolytic activity in the mucinous cysts (Fig. 1B and Fig. S1B). All 35 unique substrate cleavages that differentiated mucinous from nonmucinous cysts were enriched in the mucinous set. The degree of dysplasia within a mucinous cyst is also an important factor in determining whether surgical intervention is recommended. However, no major differences in protease activity were evident between mucinous cysts with low- or high-grade dysplasia (Fig. S2).

We generated an iceLogo frequency plot to visualize the substrate specificity pattern of the 35 mucinous-specific peptide cleavages detected at pH 3.5 (Fig. 1C) (37). At the P1 and P1′ positions, which flank the cleavage site, there was a predominant enrichment of hydrophobic amino acids with the aromatic residues tyrosine and tryptophan more favored at P1. This mirrors the previously reported substrate specificity of lysosomal aspartyl proteases (41,42).

Identification of cathepsin E and gastricsin in mucinous cysts

We next sought to identify the specific proteases within the mucinous cysts that are responsible for the increased acid-optimal cleavage of the 35 mucinous-specific substrates. To aid in the characterization of protease activity, we initially focused on a single mucinous cyst fluid sample that cleaved 30 out of the 35 substrates.

We treated the cyst fluid with broad-spectrum inhibitors against all major protease classes and analyzed changes in cleavage of the 35 mucinous-specific substrates by MSP-MS (Fig. 2A). Treatment with the aspartyl protease inhibitor pepstatin fully inhibited cleavage of 20 mucinous-specific substrates and partially inhibited cleavage of 8 additional substrates. The other broad-spectrum protease inhibitors minimally affected cleavage of the mucinous-specific substrates. The serine protease inhibitor AEBSF and the metal chelator 1,10-phenanthroline only inhibited cleavage of 3 substrates each. In a second mucinous cyst fluid sample, we confirmed that aspartyl protease inhibition with pepstatin blocks cleavage of the majority of the mucinous-specific substrates (Fig. S3).

Our inhibition data demonstrated that aspartyl proteases have increased activity in mucinous cysts. Therefore, we performed shotgun proteomic analysis of a set of mucinous (n=4) and nonmucinous cysts (n=3) to determine if there were differences in the abundance of individual aspartyl proteases. This analysis identified three aspartyl proteases – cathepsin D, cathepsin E, and gastricsin (Table S2). Label-free quantitation using precursor ion abundance, revealed that cathepsin D was present at similar levels in the mucinous and nonmucinous cysts, whereas cathepsin E and gastricsin were significantly more abundant in the mucinous cysts (Fig. 2B and Table S3). Aspartyl proteases are synthesized as inactive zymogens that undergo enzymatic maturation at an acidic pH (43). As the tumor microenvironment is known to be acidic, we investigated whether cathepsin E and gastricsin were present in the pro- or mature

forms. Exposure of fluid from a mucinous cyst to acidic pH induced a mass shift in cathepsin E and gastricsin that was comparable to that observed using recombinantly produced proteins (Fig. 2C), indicating that both proteases are released into cyst fluid in their proforms. As expected, no cathepsin E or gastricsin was detected in fluid from a representative nonmucinous cyst. Collectively, these results demonstrate that the proforms of cathepsin E and gastricsin are differentially expressed in mucinous cysts and that this induction is responsible for the increased proteolytic activity under acidic conditions.

Immunohistochemical analysis of gastricsin and cathepsin E in mucinous cysts

We further examined overexpression of cathepsin E and gastricsin in 14 mucinous cysts using immunohistochemical (IHC) analysis. Cytoplasmic gastricsin staining was observed in the epithelial cells lining 11 of the 14 mucinous cysts examined (Table S4). Interestingly, gastricsin expression was primarily associated with regions of low-grade dysplasia, and no staining was observed in regions of high-grade dysplasia (Figs. 3A-D). Gastricsin staining was also apparent in areas of low-grade dysplasia within mucinous cysts that contained both low- and high-grade dysplasia. Cytoplasmic cathepsin E was detected in all 14 mucinous cysts examined; however, staining did not show a dependence on the degree of dysplasia (Figs. 3E-H). No gastricsin or cathepsin E staining was evident in the neighboring normal ductal epithelium or stromal tissue. In addition, neither protease was detected in either of the two nonmucinous SCAs examined.

We also examined expression of gastricsin and cathepsin E in an IPMN genetic mouse model. *Ptf1a-Cre; LSL-Kras*^{G12D}*; Brg1*$^{f/f}$ mice develop cystic lesions of the pancreas that closely resemble human IPMNs (44). Consistent with the above results, we observed cytoplasmic

gastricsin and cathepsin E staining in the epithelial cells surrounding the cystic lesion (Fig. S4). Once again, there was no staining in normal pancreatic tissue.

Development of a gastricsin selective fluorescent substrate

The MSP-MS assay is ideal for discovering global differences in protease activity, but is not readily amenable for use as a diagnostic tool. Therefore, we sought to identify sensitive and selective fluorescent substrates that could be used in a standard microplate format to distinguish mucinous from nonmucinous cysts.

We focused on designing a gastricsin selective substrate as a cathepsin E selective substrate has been previously reported (45). We first analyzed the substrate specificity of recombinant cathepsin E and gastricsin using the MSP-MS assay (Fig. 4A). We also profiled recombinant cathepsin D, as it was detected in cyst fluid samples by our shotgun proteomic analysis (Fig. 2B), and therefore, we wanted to ensure that the synthesized substrates are not cleaved by this protease. Cathepsins E and D showed highly similar substrate specificity with a Pearson correlation coefficient of 0.81 and both proteases displayed a clear preference for hydrophobic residues in the P1 and P1′ positions. Gastricsin also preferred hydrophobic residues in the P1 and P1′ positions, however, direct comparison of the amino acid enrichment profiles revealed that gastricsin also has distinct cleavage preferences (Fig. 4B). Most notably, gastricsin shows a significantly stronger preference for tyrosine and tryptophan in the P1 position. Gastricsin also slightly favored small amino acids, such as glycine, serine, and alanine, in the P1' position.

Using the MSP- MS assay, we identified 75 peptides that were cleaved by gastricsin and not by cathepsins D or E (Fig. 4C). We used the specificity information from Fig. 4A-B to select

a single peptide substrate that we expected to be highly selective for gastricsin. In particular, we chose a peptide that was cleaved by gastricsin with a tryptophan and alanine in the P1 and P1′ positions, respectively. We synthesized an internally quenched, fluorescent substrate incorporating the P4 to P4′ amino acids from this peptide. This substrate was found to be greater than 120-fold selective for gastricsin over both cathepsins D and E (Fig. 4D) and is cleaved with a kcat/Km of 4.8×10^5 M^{-1}/s^{-1} (Fig. S5). We also synthesized the previously reported cathepsin E selective substrate and confirmed that it is more than 100-fold selective for cathepsin E over both cathepsin D and gastricsin (Fig. 4D) (45). Lastly, we confirmed that we could use these substrates to monitor cathepsin E and gastricsin protease activity in cyst fluid. Indeed, both substrates were cleaved in fluid from a mucinous cyst and this activity was fully inhibited by pre-incubation with pepstatin (Fig. S6).

Gastricsin and cathepsin E activity differentiate mucinous from nonmucinous cysts

We next used the gastricsin and cathepsin E fluorescent substrates to assess their relative protease activities in cyst fluid samples to determine if these activities could be used to differentiate mucinous from nonmucinous cysts. We first analyzed the 23 cyst fluid samples that we previously profiled using the MSP-MS assay. Cleavage of both the gastricsin and cathepsin E substrate was significantly higher in mucinous relative to nonmucinous cysts (Fig. S7). This prompted us to assess cathepsin E and gastricsin activity in a validation cohort comprised of an additional 87 cyst fluid samples. Again, mucinous cysts displayed significantly increased levels of gastricsin and cathepsin E activity (Fig. S7). There were no significant differences in activity between the two patient cohorts. Analysis of all 110 patient samples revealed that gastricsin activity was on average increased more than 6-fold in mucinous cysts, while cathepsin E activity

was increased only 2-fold (Fig. 5A-B). The ROC curve for gastricsin activity exhibited an area under the curve (AUC) of 0.979 for distinguishing mucinous cysts (Fig. 5C and Table S5). At the optimal cutoff of a 1.23-fold change in fluorescence, gastricsin activity demonstrated a specificity of 100% and a sensitivity of 93%. Cathepsin E activity had an AUC of 0.828 and, using this same optimal cutoff, displayed 92% specificity and 70% sensitivity for differentiating mucinous from nonmucinous cysts. Gastricsin and cathepsin E activity did not show a dependence on the type of mucinous cyst or the degree of dysplasia within a mucinous cyst (Fig. S8). Considering that gastricsin expression was only observed in regions of low-grade dysplasia (Fig. 3), we were surprised to observe that gastricsin activity was also not associated with the degree of dysplasia. This is likely because highly dysplastic and invasive mucinous lesions also often contain regions of low-grade dysplasia. We also examined whether gastricsin or cathepsin E activity were correlated with features from the revised Sendai criteria, which is a widely applied consensus guidelines for the management of mucinous cysts (8). Neither gastricsin nor cathepsin E activity showed significant differences in relation to the Sendai features we assessed (Table S6).

CEA levels were independently measured for 89 of the cyst fluid samples, and we compared abundance between mucinous (n=55) and nonmucinous cysts (n=34). As expected, CEA was significantly elevated in the mucinous cysts (Fig. S9). The CEA ROC curve exhibited an AUC of 0.865 for distinguishing mucinous cysts from nonmucinous cysts (Fig. 5C). CEA-based classification underperformed gastricsin activity, but was comparable to cathepsin E activity-based classification. For CEA, a cutoff level of 192 ng/mL is the commonly used clinical standard for differentiating mucinous from nonmucinous cysts (46). At this cutoff, CEA demonstrated a specificity of 94% and a sensitivity of 65%, which is consistent with what has

been previously reported. All 19 of the mucinous cyst fluid samples with CEA levels below the standard cutoff of 192 ng/mL were correctly classified by gastricsin activity. Additionally, the two nonmucinous cysts with CEA levels above 192 ng/mL were also correctly classified by gastricsin activity.

We also assessed whether combined analysis of CEA with gastricsin and cathepsin E activity could better differentiate mucinous from nonmucinous cysts. Gastricsin activity with CEA evaluation resulted in a classifier with an AUC of 0.998 (Fig. 5C), exhibiting a specificity of 100% and sensitivity of 98%. Inclusion of all three markers did not lead to improved differentiation of mucinous from nonmucinous cysts (Table S5).

3.6 Discussion

Although pancreatic cysts are being detected at an increasing rate, available diagnostic tests do not accurately discriminate between cyst types. Mucinous cysts have malignant potential and may require resection, while nonmucinous cysts are considered benign and require no further evaluation if these lesions are asymptomatic. Increasing the level of certainty in this distinction would spare some patients unnecessary surgical resections and reduce the need for ongoing surveillance for many more individuals. In this study, we used an unbiased and global substrate-based profiling strategy coupled with proteomics, to identify distinguishing protease activities in cyst fluid samples. Using this approach, gastricsin and cathepsin E activities were found to be promising markers for differentiating benign nonmucinous cysts from potentially malignant mucinous cysts. Selective fluorescent substrates both confirmed induction of these proteases in mucinous cysts and enabled sensitive and specific differentiation of these lesions in 110 patient samples.

To date, CEA remains the most widely used clinical biomarker for differentiating mucinous from nonmucinous cysts. However, the performance of this marker is generally considered suboptimal. Indeed, CEA analysis was only 76% accurate in our study at the standard cutoff of 192 ng/mL. Gastricsin activity was 95% accurate, and correctly classified all 21 cysts that were misclassified by CEA, clearly demonstrating the clinical utility of this marker. Furthermore, we were able to improve classification accuracy to 99% by combining CEA with gastricsin activity analysis.

Preoperatively determining the degree of dysplasia within a mucinous cyst is another major challenge for ensuring appropriate clinical intervention. However, the protease activity markers identified in this study do not differentiate between mucinous cysts with low- or high-grade dysplasia. Although this is a limitation of our markers, correctly differentiating mucinous from nonmucinous cysts is a critical first step in deciding which cysts should undergo resection. For example, pancreatic resection of the 39 benign nonmucinous cysts included in this study could potentially have been avoided through the application of our assay. In addition, 19 mucinous cysts within our patient cohort had CEA levels below the standard cutoff of 192 ng/mL. In our high-volume pancreatic centers, radiographic and clinical features allowed experienced clinicians to correctly identify these cysts as mucinous. However, medical centers without dedicated cyst specialists may be inclined to misclassify these samples as nonmucinous and would greatly benefit from our simple and accurate diagnostic assay. A number of molecular and clinical markers have recently shown promise for distinguishing mucinous cysts based on their degree of dysplasia (19,23). A sequential diagnostic strategy may emerge in which gastricsin and cathepsin E activity are used to determine if a lesion is mucinous, followed by analysis of a secondary marker to define the degree of dysplasia. Assessing gastricsin and

cathepsin E activity in combination with other promising markers will be a primary focus of future work.

Previous gene expression profiling studies of IPMNs and MCNs demonstrated overexpression of gastricsin and cathepsin E mRNA (29–31). However, the protein levels and activity of these aspartyl proteases has not been previously investigated within these lesions. Protease activity is particularly well suited to the development of a rapid and simple diagnostic test for differentiating cysts. Activity-based detection is highly sensitive because of catalytic signal amplification. Indeed, the assays described in this study use less than 5 μL of cyst fluid, whereas CEA tests often require at least 500 μL. Furthermore, unlike immunoassays, protease activity assays do not require the costly development of high-quality antibody reagents. Spectrophotometric assays can be readily adapted to the standard plate readers present in clinical laboratories, and there are already several examples of such protease activity assays in common clinical use for other indications (47,48).

We were particularly interested to observe that gastricsin expression within mucinous cysts was primarily associated with areas of low-grade dysplasia and was absent in high-grade dysplasia, although we were only able to assess four cysts containing regions of high-grade dysplasia. Previous work demonstrated that gastricsin and other foregut markers are overexpressed in other pancreatic cancer precursor lesions, reflecting a cellular dedifferentiation step prior to malignant transformation (49). Gastricsin overexpression within IPMNs and MCNs might be reflective of a similar process. In support of this hypothesis, recent work using the same IPMN genetic mouse model examined in this study showed that cellular dedifferentiation is a critical step in the development of IPMNs (44,50). Dedifferentiation within this genetic mouse model is transient and occurs prior to the development of invasive cancer, which may explain

why gastricsin expression is associated with regions of low-grade dysplasia. In contrast to gastricsin, we did not observe an association between cathepsin E expression and the degree of dysplasia present within a mucinous cyst. This suggests that different processes control the expression of these two proteases and that cathepsin E levels are less reflective of cellular identity. Additional studies using the recently developed genetic mouse models of mucinous cysts are needed to characterize how the expression of these proteases is regulated and what roles – if any – gastricsin and cathepsin E are playing in neoplastic transformation (44).

In summary, our results demonstrate that gastricsin and cathepsin E activity are sensitive and specific markers for differentiating mucinous from nonmucinous pancreatic cystic lesions. In particular, gastricsin activity is a promising candidate for the development of a simple, diagnostic test with superior performance to CEA. This could provide clinical stratification to properly manage the growing problem of pancreatic cysts.

3.7 References

1. Moris M, Bridges MD, Pooley RA, Raimondo M, Woodward TA. Association Between Advances in High-Resolution Cross-Section Imaging Technologies and Increase in Prevalence of Pancreatic Cysts From 2005 to 2014. Clin Gastroenterol Hepatol. 2016;14:585–93.

2. Lee KS, Sekhar A, Rofsky NM, Pedrosa I. Prevalence of incidental pancreatic cysts in the adult population on MR imaging. Am J Gastroenterol. 2010;105:2079–84.

3. Laffan TA, Horton KM, Klein AP, Berlanstein B, Siegelman SS, Kawamoto S, et al. Prevalence of unsuspected pancreatic cysts on MDCT. Am J Roentgenol. 2008;191:802–7.

4. Volkan Adsay N. Cystic lesions of the pancreas. Mod Pathol. 2007;20:71-93.

5. Matthaei H, Schulick RD, Hruban RH, Maitra A. Cystic precursors to invasive pancreatic cancer. Nat Rev Gastroenterol Hepatol. Nature Publishing Group; 2011;8:141–50.

6. Crippa S, Del Castillo CF, Salvia R, Finkelstein D, Bassi C, Dominguez I, et al. Mucin-Producing Neoplasms of the Pancreas: An Analysis of Distinguishing Clinical and Epidemiologic Characteristics. Clin Gastroenterol Hepatol. 2011;8:213–9.

7. Jang K-T, Park SM, Basturk O, Bagci P, Bandyopadhyay S, Stelow EB, et al. Clinicopathologic characteristics of 29 invasive carcinomas arising in 178 pancreatic mucinous cystic neoplasms with ovarian-type stroma: implications for management and prognosis. Am J Surg Pathol. 2015;39:179–87.

8. Tanaka M, Fernández-del Castillo C, Adsay V, Chari S, Falconi M, Jang J-Y, et al. International consensus guidelines 2012 for the management of IPMN and MCN of the pancreas. Pancreatology. Elsevier; 2012;12:183–97.

9.   Correa-Gallego C, Ferrone CR, Thayer SP, Wargo JA, Warshaw AL, Fernandez-Del Castillo C. Incidental pancreatic cysts: Do we really know what we are watching? Pancreatology. 2010;10:144–50.

10.   Parra-herran CE, Garcia MT, Herrera L, Bejarano PA. Cystic Lesions of the Pancreas : Clinical and Pathologic Review of Cases in a Five Year Period. J Pancreas. 2010;11:358–64.

11.   Quan SY, Visser BC, Poultsides GA, Norton JA, Chen AM, Banerjee S, et al. Predictive Factors for Surgery Among Patients with Pancreatic Cysts in the Absence of High-Risk Features for Malignancy. J Gastrointest Surg. 2015;19:1101–5.

12.   Cho CS, Russ AJ, Loeffler AG, Rettammel RJ, Oudheusden G, Winslow ER, et al. Preoperative classification of pancreatic cystic neoplasms: the clinical significance of diagnostic inaccuracy. Ann Surg Oncol. 2013;20:3112–9.

13.   Park WG, Mascarenhas R, Palaez-Luna M, Smyrk TC, Kane DO, Ph D, et al. Diagnostic Performance Of Cyst Fluid Carcinoembryonic Antigen And Amylase In Histologically Confirmed Pancreatic Cysts. Pancreas. 2011;40:42–5.

14.   Ngamruengphong S, Bartel MJ, Raimondo M. Cyst carcinoembryonic antigen in differentiating pancreatic cysts : A. Dig Liver Dis. Editrice Gastroenterologica Italiana; 2013;45:920–6.

15.   Almoguera C, Shibata D, Forrester K, Martin J, Arnheim N, Perucho M. Most Human Carcinomas of the Exocrine Contain Mutant c-K-ras Genes. Cell. 1988;53:549–54.

16.   Hezel AF, Kimmelman AC, Stanger BZ, Bardeesy N, Depinho R a. Genetics and biology of pancreatic ductal adenocarcinoma. Genes Dev. 2006;20:1218–49.

17.   Khalid A, Zahid M, Finkelstein SD, Leblanc JK. Pancreatic cyst fluid DNA analysis in

evaluating pancreatic cysts : a report of the PANDA study. Gastrointest Endosc. 2009;69:1095–102.

18. Wu J, Matthaei H, Maitra A, Dal Molin M, Wood LD, Eshleman JR, et al. Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development. Sci Transl Med. 2011;3:92ra66.

19. H T. Hata, M. Dal Molin, M. Suenaga, J. Yu, M. Pittman, M. Weiss, M. I. Canto, C. Wolfgang, A. M. Lennon, R. H. Hruban, M. Goggins. Cyst fluid telomerase activity predicts the histologic grade of cystic neoplasms of the pancreas. Clin Cancer Res. 2016;20:5141-41.

20. Zikos T, Pham K, Bowen R, Chen AM, Banerjee S, Friedland S, et al. Cyst Fluid Glucose is Rapidly Feasible and Accurate in Diagnosing Mucinous Pancreatic Cysts. Am J Gastroenterol. 2015;110:909–14.

21. Yip-Schneider MT, Wu H, Dumas RP, Hancock BA. Vascular Endothelial Growth Factor , a Novel and Highly Accurate Pancreatic Fluid Biomarker for Serous Pancreatic Cysts. J Am Coll Surg. 2014;218:608–17.

22. Cao Z, Maupin K, Curnutte B, Fallon B, Feasley CL, Brouhard E, et al. Specific glycoforms of MUC5AC and endorepellin accurately distinguish mucinous from nonmucinous pancreatic cysts. Mol Cell Proteomics. 2013;12:2724–34.

23. Maker A V, Katabi N, Qin L, Klimstra DS, Schattner M, Brennan F, et al. Cyst fluid interleukin-1b (IL1b) levels predict the risk of carcinoma in intraductal papillary mucinous neoplasms of the pancreas. Clin Cancer Res. 2012;17:1502–8.

24. Kessenbrock K, Plaks V, Werb Z. Matrix metalloproteinases: regulators of the tumor microenvironment. Cell. 2010;141:52–67.

25. Sevenich L, Joyce JA. Pericellular proteolysis in cancer. Genes Dev. 2014;2331–47.

26. Sobotič B, Vizovišek M, Vidmar R, Van Damme P, Gocheva V, Joyce JA, et al. Proteomic Identification of Cysteine Cathepsin Substrates Shed from the Surface of Cancer Cells. Mol Cell Proteomics. 2015;14:2213–28.

27. Joyce J a, Baruch A, Chehade K, Meyer-Morse N, Giraudo E, Tsai F-Y, et al. Cathepsin cysteine proteases are effectors of invasive growth and angiogenesis during multistage tumorigenesis. Cancer Cell. 2004;5:443–53.

28. Gocheva V, Zeng W, Ke D, Klimstra D, Reinheckel T, Peters C, et al. Distinct roles for cysteine cathepsin genes in multistage tumorigenesis. Genes Dev. 2006;543–56.

29. Terris B, Blaveri E, Crnogorac-Jurcevic T, Jones M, Missiaglia E, Ruszniewski P, et al. Characterization of gene expression profiles in intraductal papillary-mucinous tumors of the pancreas. Am J Pathol. 2002;160:1745–54.

30. Sato N, Fukushima N, Maitra A, Iacobuzio-Donahue C a, van Heek NT, Cameron JL, et al. Gene expression profiling identifies genes associated with invasive intraductal papillary mucinous neoplasms of the pancreas. Am J Pathol. 2004;164:903–14.

31. Fukushima N, Sato N, Prasad N, Leach SD, Hruban RH, Goggins M. Characterization of gene expression in mucinous cystic neoplasms of the pancreas using oligonucleotide microarrays. Oncogene. 2004;23:9042–51.

32. Ke E, Patel BB, Liu T, Li X, Haluszka O, Hoffman JP, et al. Proteomic Analyses of Pancreatic Cyst Fluids. Pancreas. 2009;38:1–21.

33. Räty S, Sand J, Laukkarinen J, Vasama K, Bassi C, Salvia R, et al. Cyst fluid SPINK1 may help to differentiate benign and potentially malignant cystic pancreatic lesions. Pancreatology; 2013;13:530–3.

34.   O'Donoghue AJ, Eroy-reveles AA, Knudsen GM, Ingram J, Zhou M, Statnekov JB, et al. Global identification of peptidase specificity by multiplex substrate profiling. Nat Methods. 2012;9:1095-100.

35.   Chalkley RJ, Baker PR, Medzihradszky KF, Lynn AJ, Burlingame AL. In-depth Analysis of Tandem Mass Spectrometry Data from Disparate Instrument Types. Mol Cell Proteomics. 2008;2386–98.

36.   Schilling B, Rardin MJ, MacLean BX, Zawadzka a. M, Frewen BE, Cusack MP, et al. Platform-Independent And Label-Free Quantitation Of Proteomic Data Using MS1 Extracted Ion Chromatograms In Skyline: Application To Protein Acetylation And Phosphorylation. Mol Cell Proteomics. 2012;11:202–14.

37.   Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K. Improved visualization of protein consensus sequences by iceLogo. Nat Methods. 2009;6:786–7.

38.   Winter MB, Salcedo EC, Lohse MB, Hartooni N, Gulati M, Sanchez H, et al. Global Identification of Biofilm-Specific Proteolysis in Candida albicans. MBio. 2016;7:1–13.

39.   O'Donoghue AJ, Knudsen GM, Beekman C, Perry J a., Johnson AD, DeRisi JL, et al. Destructin-1 is a collagen-degrading endopeptidase secreted by Pseudogymnoascus destructans, the causative agent of white-nose syndrome. Proc Natl Acad Sci. 2015;112:7478-83

40.   Small JL, O'Donoghue AJ, Boritsch EC, Tsodikov O V, Knudsen GM, Vandal O, et al. Substrate specificity of MarP, a periplasmic protease required for resistance to acid and oxidative stress in Mycobacterium tuberculosis. J Biol Chem. 2013;288:12489–99.

41.   Impens F, Colaert N, Helsens K, Ghesquière B, Timmerman E, De Bock P-J, et al. A quantitative proteomics design for systematic identification of protease cleavage events.

Mol Cell Proteomics. 2010;9:2327–33.

42. Donoghue AJO, Ivry SL, Chaudhury C, Hostetter DR, Hanahan D, Craik CS. Procathepsin E is highly abundant but minimally active in pancreatic ductal adenocarcinoma tumors. Biol Chem. 2016;397:871–81.

43. Dunn BM. Structure and Mechanism of the Pepsin-Like Family of Aspartic Peptidases. Chem Rev. 2002;102:4431–58.

44. von Figura G, Fukuda A, Roy N, Liku ME, Morris JP, Kim GE, et al. The chromatin regulator Brg1 suppresses formation of intraductal papillary mucinous neoplasm and pancreatic ductal adenocarcinoma. Nat Cell Biol. 2014;16:255–67.

45. Abd-Elgaliel WR, Tung C-H. Selective detection of Cathepsin E proteolytic activity. Biochim Biophys Acta. 2010;1800:1002–8.

46. Brugge WR, Lewandrowski K, Lee-Lewandrowski E, Centeno BA, Szydlo T, Regan S, et al. Diagnosis of Pancreatic Cystic Neoplasms: A Report of the Cooperative Pancreatic Cyst Study. Gastroenterology. 2004;126:1330–6.

47. Kremer Hovinga JA, Mottini M, Lämmle B. Measurement of ADAMTS-13 activity in plasma by the FRETS-VWF73 assay: Comparison with other assay methods. J Thromb Haemost. 2006;4:1146–8.

48. Moll S, Ortel TL. Monitoring Warfarin Therapy in Patients with Lupus Anticoagulants. Ann Intern Med. 1997;127:177–85.

49. Prasad NB, Biankin A V, Fukushima N, Maitra A, Dhara S, Elkahloun AG, et al. Gene Expression Profiles in Pancreatic Intraepithelial Neoplasia Reflect the Effects of Hedgehog Signaling on Pancreatic Ductal Epithelial Cells Gene Expression Profiles in Pancreatic Intraepithelial Neoplasia Reflect the Effects of Hedgehog Signaling on. Cancer

Res. 2005;1619–26.

50.   Roy N, Malik S, Villanueva KE, Urano A, Lu X, Von Figura G, et al. Brg1 promotes both tumor-suppressive and oncogenic activities at distinct stages of pancreatic cancer formation. Genes Dev. 2015;29:658–71.

3.8 Figures

Figure 1

Figure 3.1 Comparison of global proteolytic activity in mucinous and nonmucinous cysts by MSP-MS. Volcano plots displaying the peptide cleavages generated by mucinous (n=16) and nonmucinous cysts (n=7) when assayed at pH 7.5 **(A)** or pH 3.5 **(B)**. Spectral counts of peptide cleavage products were used for relative quantification of the fold change (mucinous/nonmucinous) and hypothesis testing. Cleavages that met the criteria for differentiating mucinous from nonmucinous cysts (+/- 1 $\log_2$(fold change), $p < 0.05$) are shown in blue. The substrate specificity of the cleavages within the red box is displayed with an iceLogo plot **(C)**. Residues shown are statistically significant with $p < 0.05$.

Figure 2
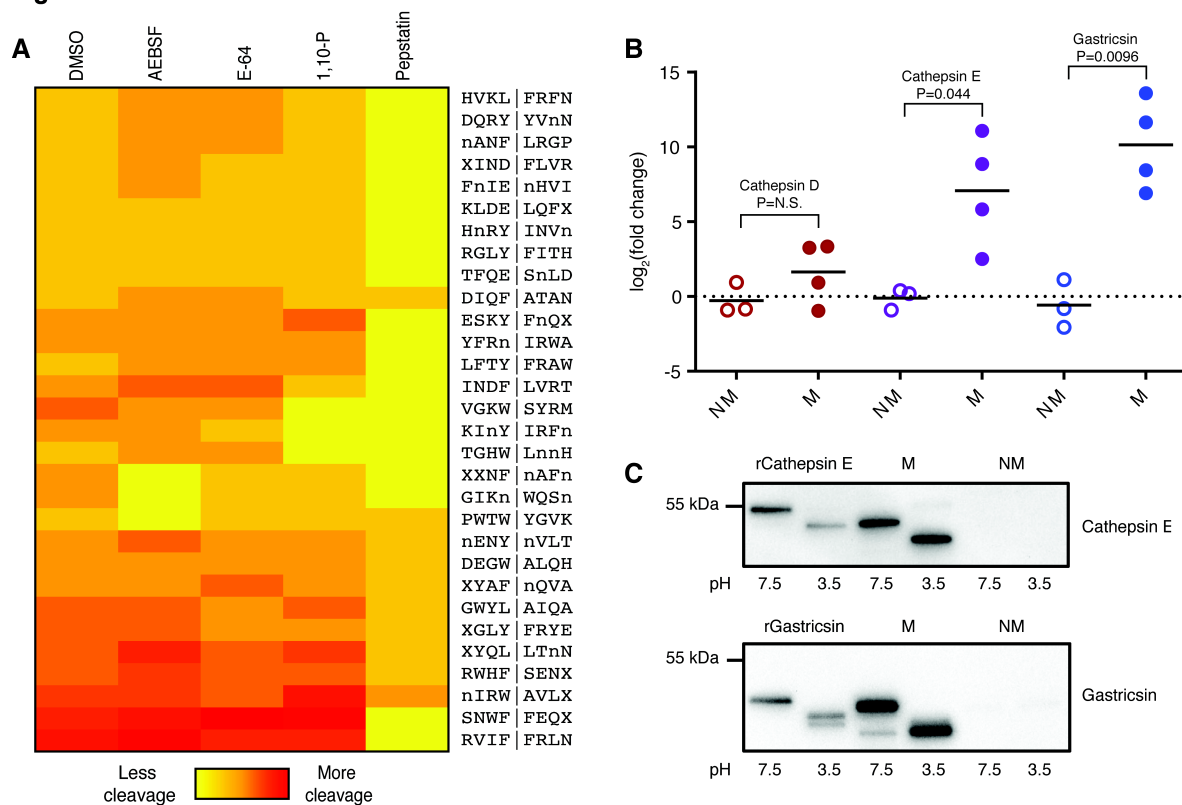
Figure 3.2 Identification of enriched aspartyl protease activity in mucinous cysts. (A) Heatmap displaying cleavage of 30 mucinous-specific substrates following treatment of a mucinous cyst fluid sample with DMSO or various broad-spectrum protease inhibitors. Spectral counts were used for relative quantification of peptide cleavage products. Vertical bar (|) indicates the site of cleavage within substrates. (B) Label-free quantitation of aspartyl protease relative abundance in mucinous (M) and nonmucinous (NM) cysts. (C) Western blot analysis of recombinant (r) and cyst fluid-derived cathepsin E and gastricsin. Samples were pre-incubated at the indicated pH for 10 minutes.
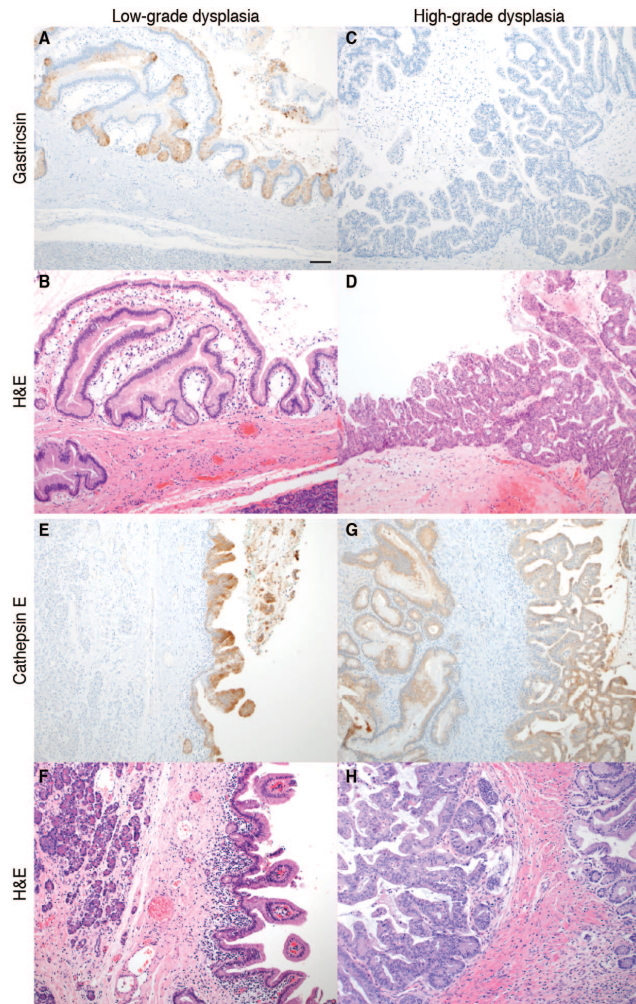
Figure 3.3 Immunohistochemical analysis of gastricsin and cathepsin E in mucinous cysts. Histological analysis of mucinous cysts with low-grade dysplasia (**A, B, E, F**) and high-grade dysplasia (**C, D, G, H**). Gastricsin (**A, C**), cathepsin E **(E, G)**, and haematoxylin and eosin (H&E) staining (**B, D, F, H**) in IPMNs **(A-F)** and MCNs **(G, H)**. Scale bar is 10 μm.
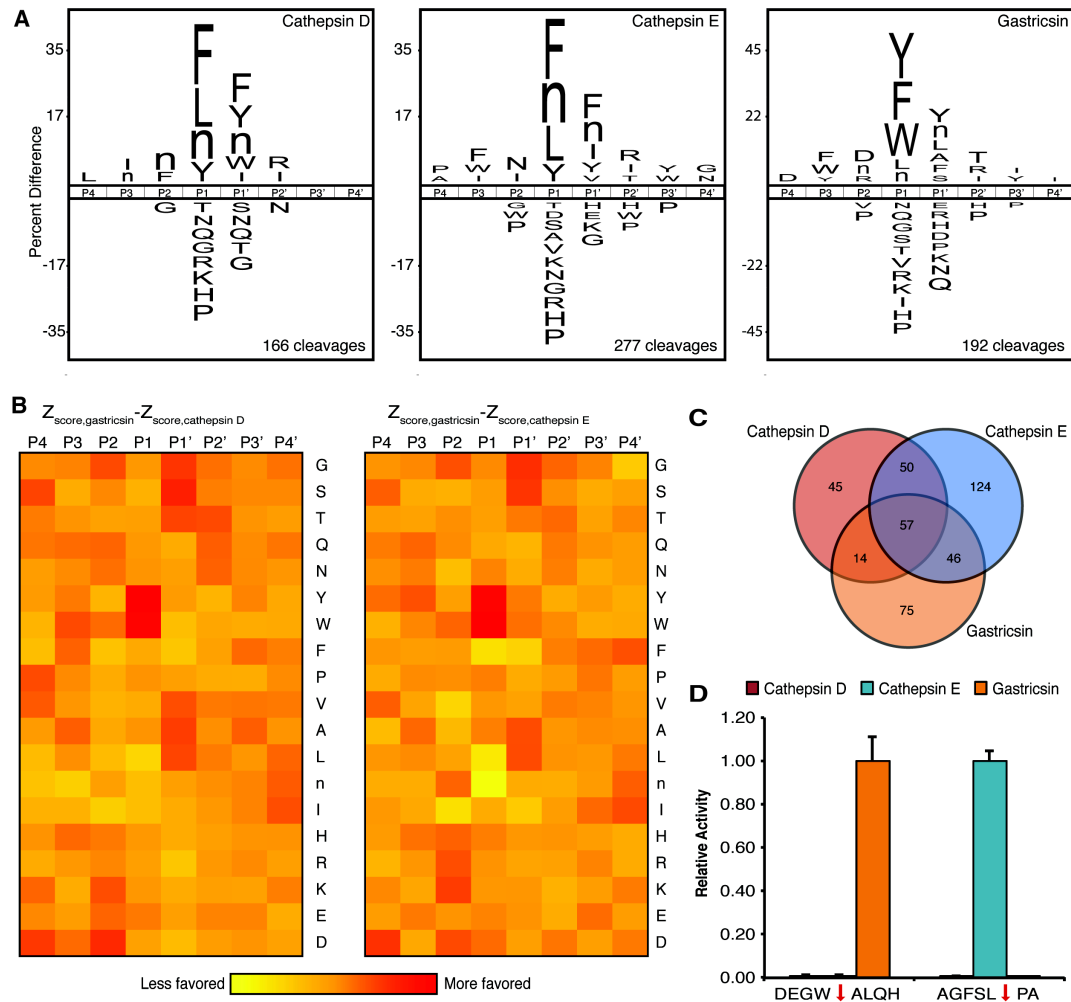
Figure 4

Figure 3.4 Design and synthesis of gastricsin selective fluorescent substrate. (A) Substrate specificity of cathepsin D, cathepsin E, and gastricsin as determined by MSP-MS. Residues shown in iceLogo are statistically significant with $p < 0.05$. **(B)** Heatmap comparing the amino acid enrichment Z-scores for gastricsin relative to cathepsin D and cathepsin E. **(C)** Venn diagram depicting the unique and overlapping cleavages detected by MSP-MS with cathepsin D, cathepsin E, and gastricsin. **(D)** Cleavage of the fluorescent substrates by cathepsin D, cathepsin E, and gastricsin. Activity was normalized to 1.00 based on the maximal activity against each substrate. Red arrow indicates the site of cleavage. Error bars denote standard error of the mean (SEM) from triplicate analysis.

100

Figure 5
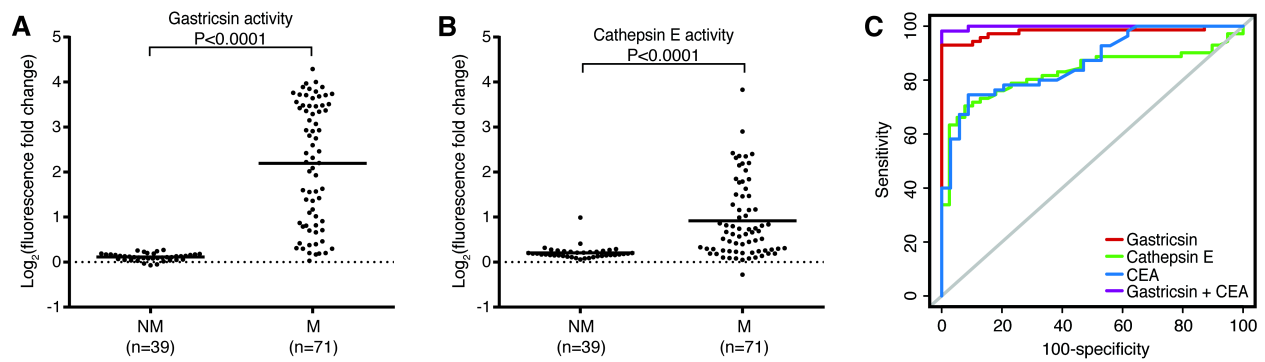
Fig. 3.5 Quantification of gastricsin and cathepsin E activity in 110 cyst fluid samples.
Analysis of gastricsin **(A)** and cathepsin E **(B)** activity in nonmucinous (NM) and mucinous (M) cysts using fluorescent substrates. **(C)** ROC curves comparing sensitivity and specificity of CEA, gastricsin, cathepsin E, and CEA and gastricsin in combination.
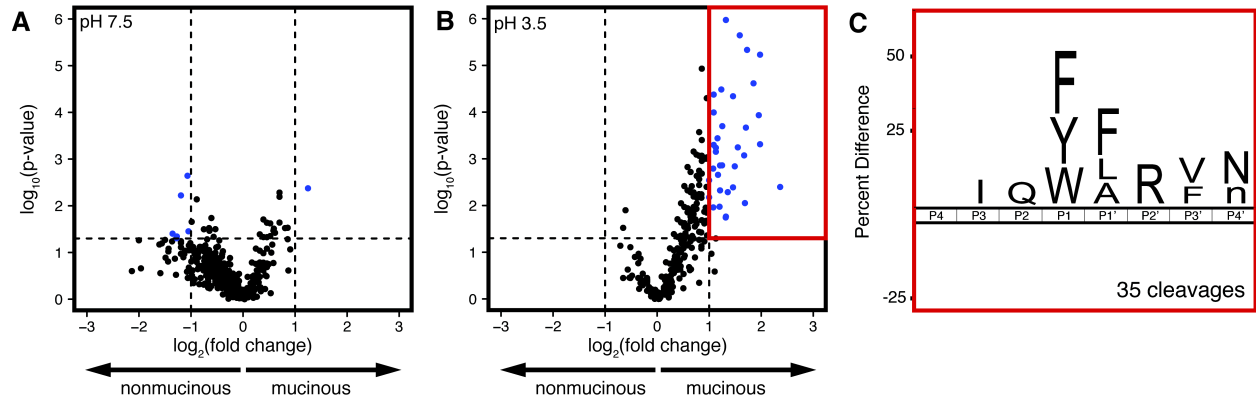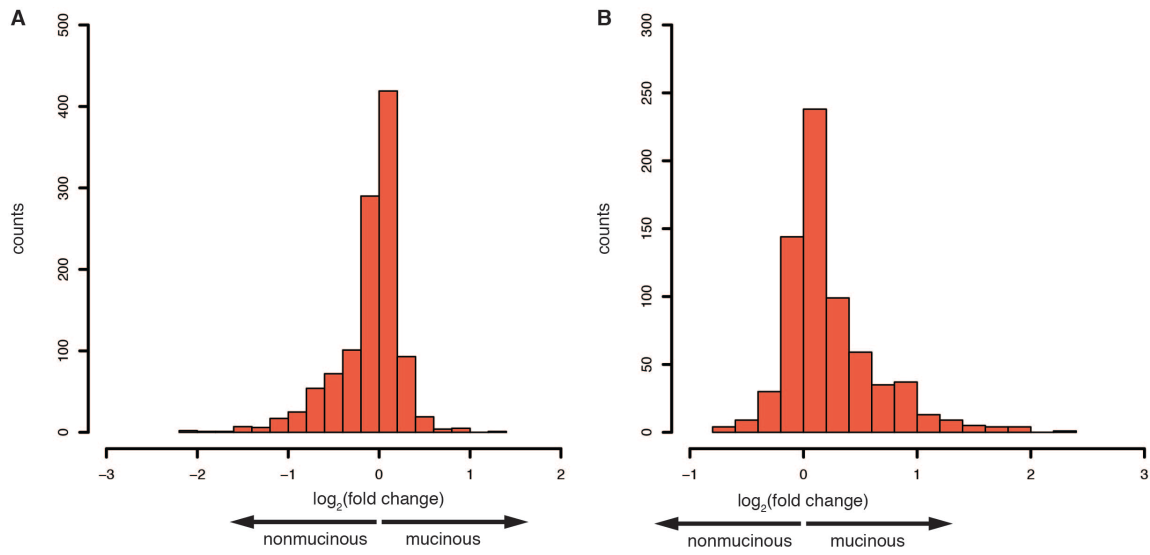
Supplementary Figure 3.1 Comparison of global proteolytic activity in mucinous and nonmucinous cysts by MSP-MS. Histograms depicting the number of cleavages enriched in mucinous and nonmucinous cysts at pH 3.5 **(B)** and pH 7.5 **(B)**. Spectral counts of peptide cleavage products were used for quantification of the fold change (mucinous/nonmucinous) and counts are the number of peptide cleavages in each bin.
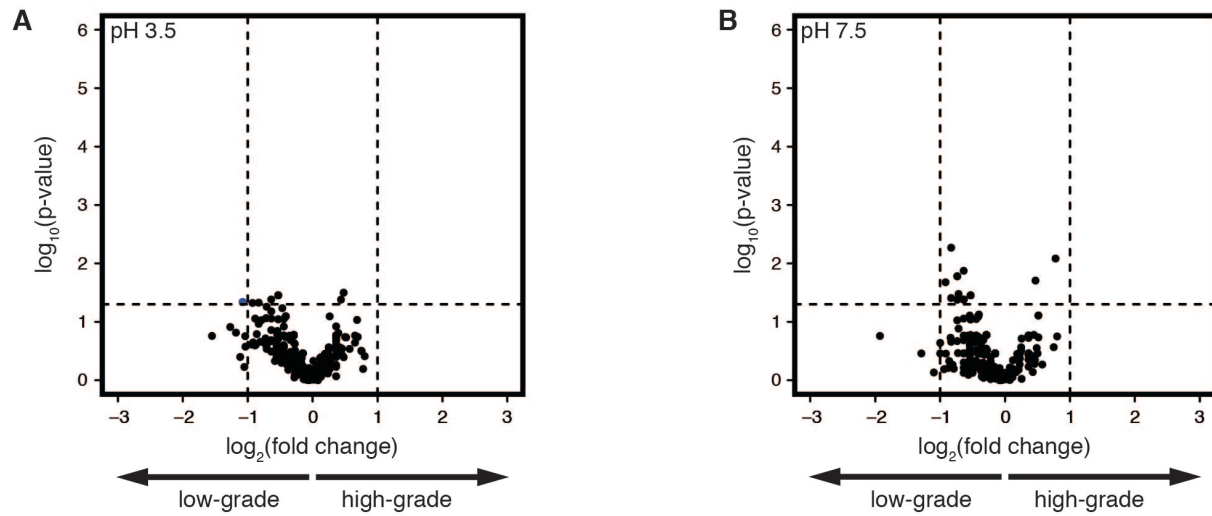
Supplementary Figure 3.2 Comparison of global proteolytic activity in mucinous cysts with low- and high-grade dysplasia. Volcano plots displaying the peptide cleavages generated by the mucinous cysts with low-grade dysplasia (n=9) and high-grade dysplasia (n=7) at pH 3.5 **(A)** and pH 7.5 **(B)**. Fold change corresponds to low-grade dysplasia/high-grade dysplasia.

DMSO Pepstatin

XYAF | nQVA
XYQL | LTnN
nANF | LRGP
SNWF | FEQX
INDF | LVRT
GWYL | AIQA
YFRn | IRWA
LFTY | FRAW
nENY | nVLT
DQRY | YVnN
HVKL | FRFN
ESKY | FnQX
KIMY | IRFN
FnIE | nHVI
DIQF | ATAN
TFQE | SnLD
XPDF | YLGR
KLDE | LQFX
XIND | FLVR
GIKn | WQSn
RGLY | FITH
RVIF | FRLN
EYQF | VTYK

Less cleavage — More cleavage
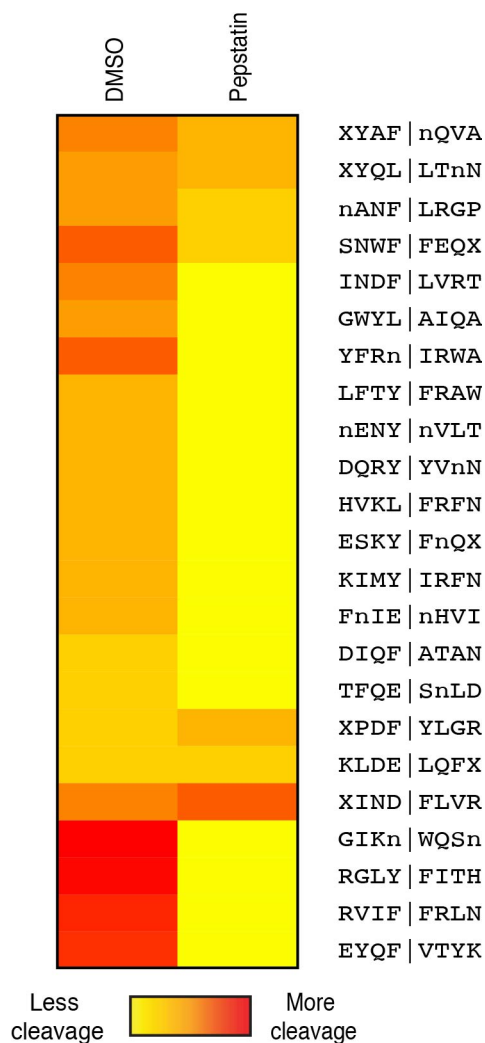
Supplementary Figure 3.3 Analysis of pepstatin inhibition on protease activity through MSP-MS. Heatmap displaying cleavage of 28 mucinous-specific substrates following treatment of a mucinous cyst fluid sample with DMSO or pepstatin. Spectral counts were used for relative quantification of peptide cleavage products. Vertical bar (|) indicates the site of cleavage within substrates.
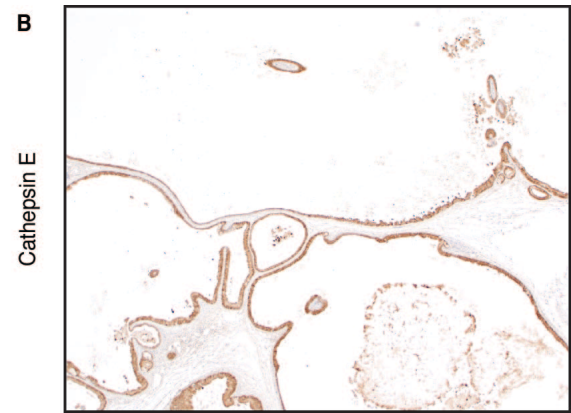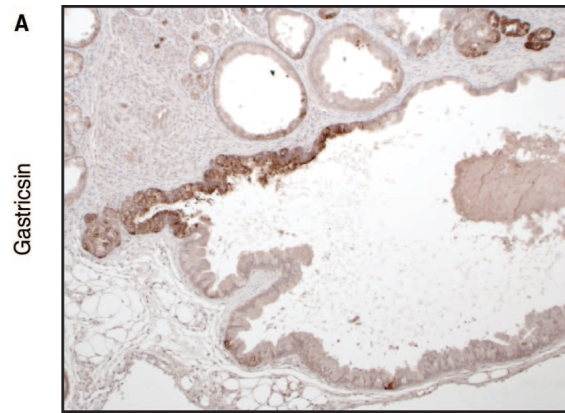
Supplementary Figure 3.4 Immunohistochemical analysis of IPMN genetic mouse model. Immunohistochemical analysis of gastricsin **(A)** and cathepsin E **(B)** in a cystic lesion from a 40-week-old *Ptf1a-Cre; LSL-Kras^{G12D}; Brg1$^{f/f}$* mouse.
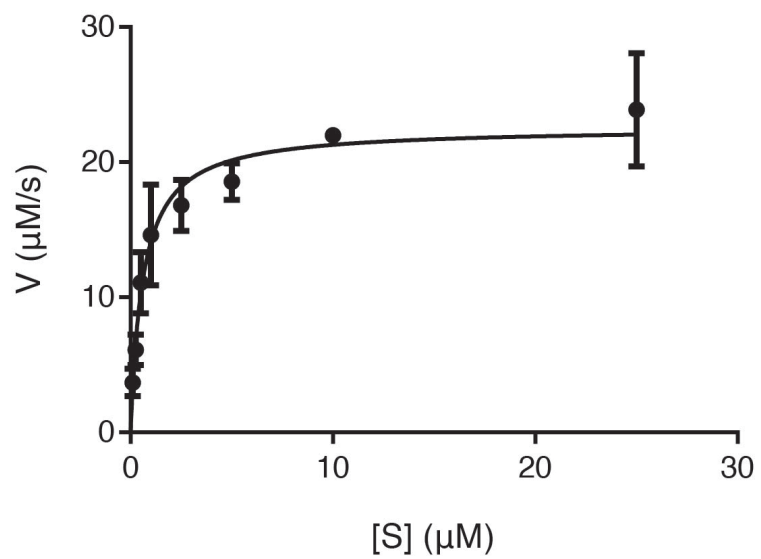
Supplementary Figure 3.5 Kinetic analysis of gastricsin selective substrate. Michaelis-Menten analysis of cleavage of DEGW | ALQH substrate by gastricsin. kcat/Km = X, Vmax is Y. Error bars denote SEM from triplicate analysis.
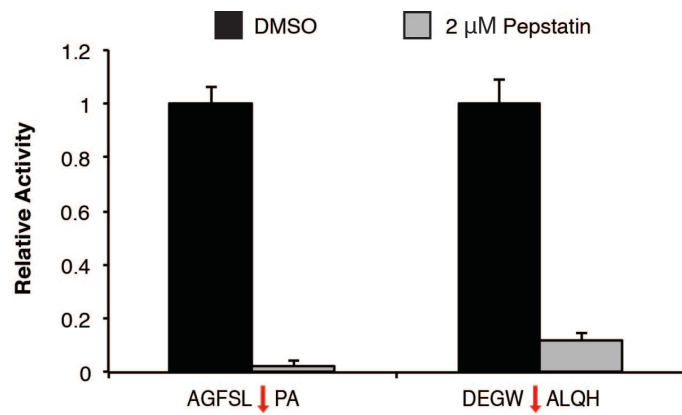
Supplementary Figure 3.6 Cleavage of aspartyl protease substrates in mucinous cysts. Pepstatin inhibition of cleavage of fluorescent substrates by a mucinous cyst fluid sample. Activity was normalized relative to DMSO control treatment and error bars denote SEM from triplicate analysis.
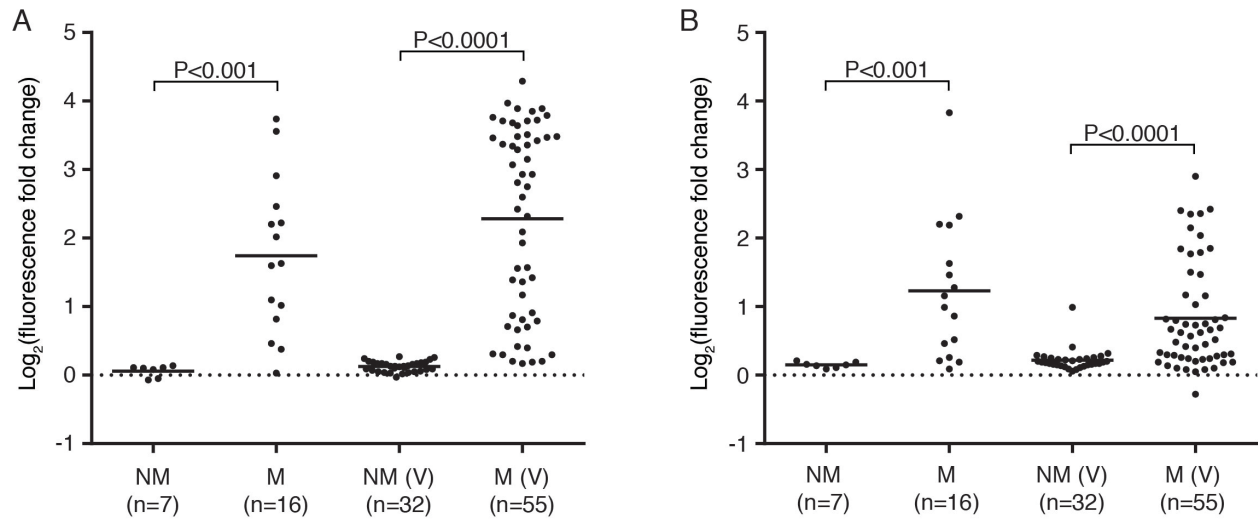
Supplementary Figure 3.7 Quantification of aspartyl protease activity in cyst fluid samples using fluorescent peptide substrates. Gastricsin **(A)** and cathepsin E **(B)** activity in samples analyzed by MSP-MS and fluorescence (n=23) and in samples from validation (V) cohort (n=87) that were just assayed using fluorescent substrates.
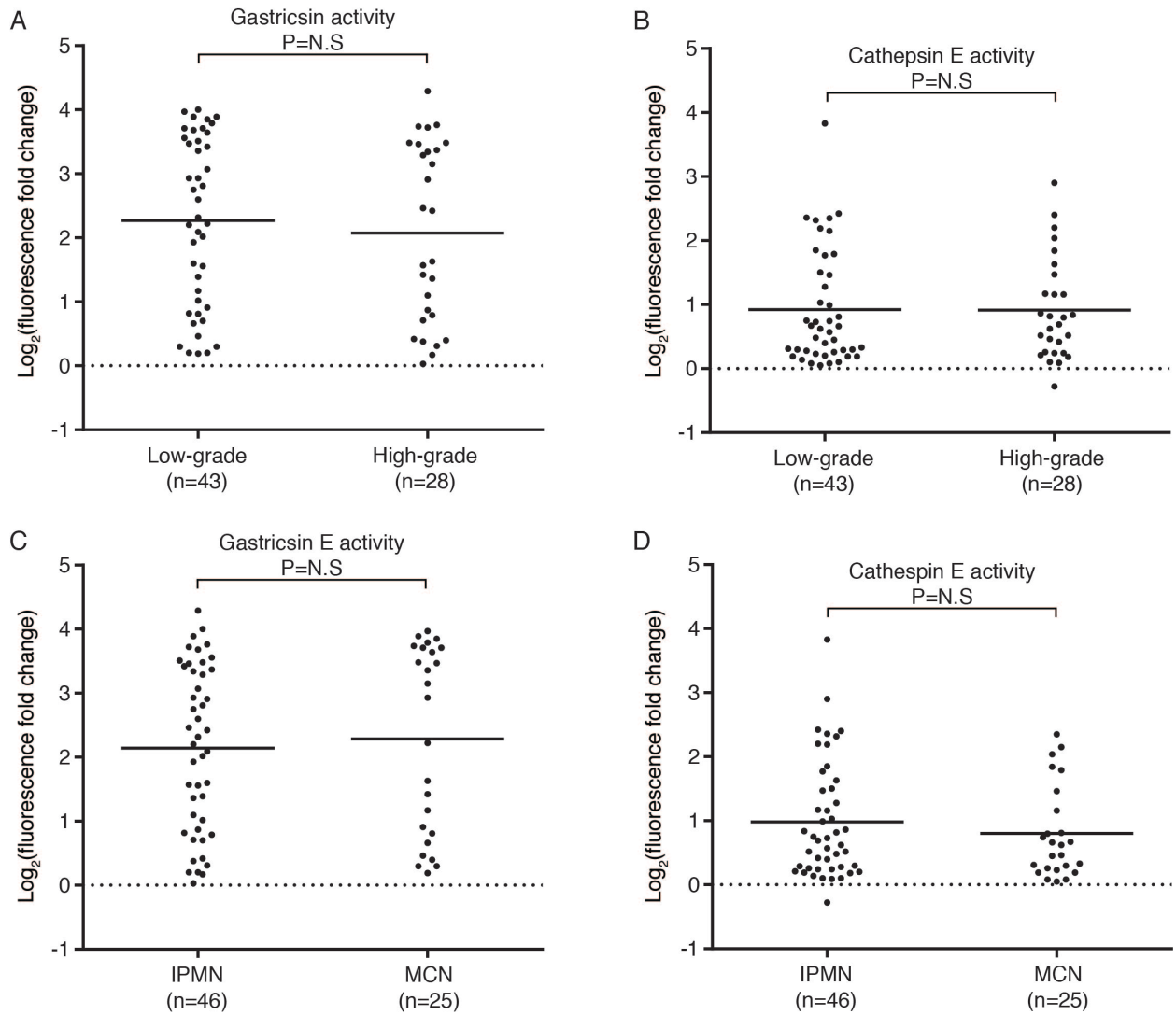
Supplementary Figure 3.8 Analysis of gastricsin and cathepsin E activity in 71 mucinous cyst fluid samples. Gastricsin and cathepsin E activity in mucinous cysts with low- and high-grade dysplasia **(A, B)** or IPMNs and MCNs **(C, D)**.
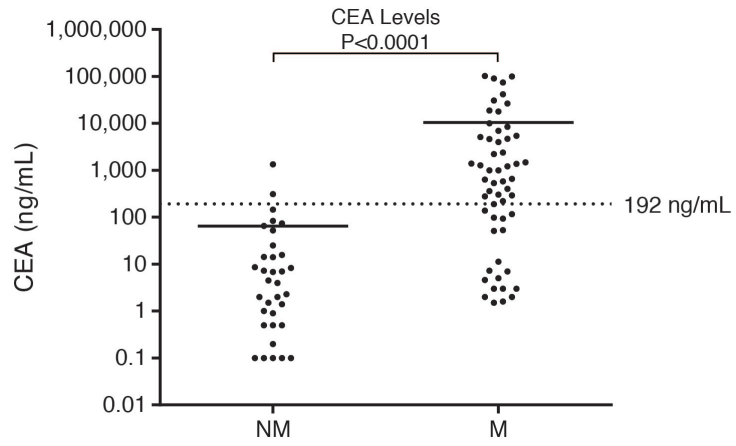
Supplementary Figure 3.9 CEA levels in mucinous and nonmucinous cysts. Dashed line indicates the standard clinical cutoff of 192 ng/mL.

Tables

Supplementary Table 3.1 Characteristics of 110 patients analyzed in this study.

Internal ID	Diagnosis	Highest grade of dysplasia	Cyst size (mm)	Collection method (EUS/Surgery)	Age	Gender	Institution
14	IPMN	High	22	Surgery	60	Female	UCSF
15	IPMN	High	19	Surgery	69	Male	UCSF
23	IPMN	Low	8	Surgery	55	Female	UCSF
34	IPMN	High	59	Surgery	78	Male	UCSF
55	IPMN	High	26	Surgery	69	Female	Stanford
138	IPMN	Low	65	EUS	68	Male	Stanford
770	IPMN	High	39	Surgery	72	Female	Stanford
775	IPMN	High	120	Surgery	78	Male	Stanford
788	IPMN	High	18	Surgery	64	Male	Stanford
797	IPMN	High	30	Surgery	79	Female	Stanford
801	IPMN	High	25	Surgery	68	Male	Stanford
41	IPMN	Low	18	Surgery	41	Male	Pittsburgh
47	IPMN	Moderate	31	Surgery	47	Female	Pittsburgh
51	IPMN	High	20	Surgery	51	Male	Pittsburgh
52	IPMN	Moderate	47	Surgery	52	Female	Pittsburgh
53	IPMN	Low	25	Surgery	53	Female	Pittsburgh
54	IPMN	Moderate	38	Surgery	54	Female	Pittsburgh
62	IPMN	Moderate	40	Surgery	62	Female	Pittsburgh
64	IPMN	Moderate	47	Surgery	64	Male	Pittsburgh
69	IPMN	Moderate	30	Surgery	69	Female	Pittsburgh
77	IPMN	Moderate	65	Surgery	77	Female	Pittsburgh
85	IPMN	High	27	Surgery	85	Female	Pittsburgh
90	IPMN	High	NA	EUS	69	Female	Indiana
169	IPMN	High	35	ERCP	75	Male	Indiana
1177	IPMN	High	57	Surgery	67	Female	Indiana
1183	IPMN	High	120	Surgery	78	Male	Indiana
1187	IPMN	High	NA	Surgery	66	Female	Indiana
1209	IPMN	High	50	Surgery	74	Male	Indiana
1217	IPMN	High	23	Surgery	58	Male	Indiana
1233	IPMN	High	NA	Surgery	71	Female	Indiana
1250	IPMN	Low	23	Surgery	68	Female	Indiana
1251	IPMN	Low	25	Surgery	79	Male	Indiana
1252	IPMN	High	11	Surgery	72	Male	Indiana
1253	IPMN	High	30	Surgery	63	Female	Indiana
1255	IPMN	Low	37	Surgery	80	Female	Indiana
1272	IPMN	Low	20	Surgery	40	Female	Indiana
1275	IPMN	High	20	Surgery	62	Male	Indiana
157	IPMN	Low	35	Surgery	84	Male	Stanford
685	IPMN	Low	4	Surgery	71	Female	Stanford
733	IPMN	Low	45	EUS	78	Male	Stanford
755	IPMN	Moderate	60	Surgery	60	Male	Stanford
760	IPMN	Moderate	35	Surgery	84	Female	Stanford
761	IPMN	Moderate	30	Surgery	74	Female	Stanford
785	IPMN	Low	90	Surgery	72	Male	Stanford
786	IPMN	Moderate	50	Surgery	71	Female	Stanford
4	MCN	Moderate	23	Surgery	57	Female	UCSF
6	MCN	High	22	Surgery	74	Female	UCSF
125	MCN	High	57	Surgery	54	Female	UCSF
130	MCN	High	125	Surgery	47	Female	UCSF
30	MCN	Moderate	60	Surgery	48	Male	Stanford
136	MCN	Low	100	Surgery	44	Female	Stanford
139	MCN	Low	140	Surgery	22	Female	Stanford

Supplementary Table 3.1 (continued)

Internal ID	Diagnosis	Highest grade of dysplasia	Cyst size (mm)	Collection method (EUS/Surgery)	Age	Gender	Institution
686	MCN	Low	125	Surgery	44	Female	Stanford
716	MCN	Low	30	Surgery	65	Female	Stanford
720	MCN	Low	45	Surgery	49	Female	Stanford
752	MCN	High	152	Surgery	42	Female	Stanford
768	MCN	Low	65	Surgery	25	Female	Stanford
787	MCN	Low	30	Surgery	76	Female	Stanford
799	MCN	Moderate	70	Surgery	31	Female	Stanford
42	MCN	Moderate	100	Surgery	42	Female	Pittsburgh
50	MCN	High	215	Surgery	50	Female	Pittsburgh
59	MCN	Moderate	19	Surgery	59	Female	Pittsburgh
74	MCN	Low	115	Surgery	74	Female	Pittsburgh
75	MCN	Low	30	Surgery	75	Female	Pittsburgh
1152	MCN	Low	18	Surgery	61	Female	Indiana
1182	MCN	Low	50	Surgery	25	Female	Indiana
1191	MCN	Low	25	Surgery	24	Female	Indiana
1261	MCN	Low	40	Surgery	24	Female	Indiana
1263	MCN	High	28	Surgery	44	Male	Indiana
1265	MCN	Low	30	Surgery	50	Female	Indiana
1269	MCN	Low	50	Surgery	57	Female	Indiana
81	Pseudocyst	NA	55	EUS	48	Male	Stanford
98	Pseudocyst	NA	219	Surgery	30	Female	Stanford
101	Pseudocyst	NA	40	EUS	62	Female	Stanford
126	Pseudocyst	NA	90	EUS	57	Male	Stanford
151	Pseudocyst	NA	12	Surgery	55	Female	Stanford
678	Pseudocyst	NA	100	Surgery	56	Female	Stanford
705	Pseudocyst	NA	90	Surgery	60	Male	Stanford
43	Pseudocyst	NA	79	EUS	43	Male	Pittsburgh
49	Pseudocyst	NA	98	EUS	49	Male	UCSF
63	Pseudocyst	NA	91	EUS	63	Male	Pittsburgh
70	Pseudocyst	NA	100	Surgery	70	Male	Pittsburgh
78	Pseudocyst	NA	31	EUS	78	Female	Pittsburgh
11	Pseudocyst	NA	NA	EUS	68	Female	Indiana
50	Pseudocyst	NA	NA	Surgery	59	Female	Indiana
159	Pseudocyst	NA	NA	EUS	67	Male	Indiana
311	Pseudocyst	NA	NA	EUS	60	Male	Indiana
1142	Pseudocyst	NA	NA	Surgery	52	Male	Indiana
1178	Pseudocyst	NA	NA	Surgery	66	Female	Indiana
1180	Pseudocyst	NA	NA	Surgery	42	Female	Indiana
1191	Pseudocyst	NA	NA	Surgery	24	Female	Indiana
1230	Pseudocyst	NA	NA	Surgery	58	Female	Indiana
21	SCA	NA	20	Surgery	58	Male	UCSF
87	SCA	NA	160	Surgery	62	Male	UCSF
68	SCA	NA	50	EUS	59	Male	Stanford
147	SCA	NA	38	EUS	45	Female	Stanford
679	SCA	NA	65	Surgery	70	Female	Stanford
732	SCA	NA	100	Surgery	43	Female	Stanford
748	SCA	NA	60	Surgery	66	Male	Stanford
767	SCA	NA	105	Surgery	47	Male	Stanford
802	SCA	NA	60	Surgery	75	Female	Stanford
10	SCA	NA	40	Surgery	55	Male	Indiana
121	SCA	NA	110	EUS	33	Female	Indiana
128	SCA	NA	25	EUS	55	Female	Indiana
409	SCA	NA	45	Surgery	26	Female	Indiana
852	SCA	NA	25	Surgery	30	Female	Indiana
1053	SCA	NA	12	Surgery	41	Female	Indiana
1218	SCA	NA	82	Surgery	76	Female	Indiana
1234	SCA	NA	93	Surgery	76	Male	Indiana
1283	SCA	NA	30	Surgery	74	Female	Indiana

Supplementary Table 3.4 Cathepsin E and gastricsin staining in cyst tissue.

Protein	IPMN (HGD)	IPMN (LGD)	MCN (HGD)	MCN (LGD)	SCA
Gastricsin	0/3	4/4	0/1	4/6	0/2
Cathepsin E	3/3	4/4	1/1	6/6	0/2

Supplementary Table 3.5 Diagnostic performance of individial and combination markers in cyst fluid samples.

Markers	AUC (95% CI)	Sensitivity (%) (95% CI)	Sensitivity (%) (95% CI)
Gastricsin + Cathepsin E + CEA	0.998 (0.995 - 1)	98.2 (90.3-100)	100 (89.7-100)
Gastricsin + CEA	0.998 (0.995 - 1)	98.2 (90.3-100)	100 (89.7-100)
Gastricsin + Cathepsin E	0.987 (0.971 - 1)	94.4 (0.9713 - 1)	100 (91.0-100)
Cathepsin E + CEA	0.909 (0.845 - 0.973)	83.6 (71.2-92.2)	94.1 (80.3-99.3)
Gastricsin	0.979 (0.952 - 1)	93.0 (84.3-97.7)	100 (91.0-100)
Cathepsin E	0.828 (0.75 - 0.907)	70.4 (58.4-80.7)	92.3 (79.1-98.4)
CEA	0.865 (0.792 - 0.938)	74.5 (61.5-85.3)	91.2 (76.3-98.14)

Supplementary Table 3.6 Protease activity data in relation to revised Sendai criteria for 71 mucinous cysts.

Mucinous cyst features	Fraction of mucinous cysts	Gastricsin activity: fluorescence fold change	Cathepsin E activity: fluorescence fold change
All mucinous cysts	71/71	6.56 (5.11)	2.31 (2.00)
Worrisome features	56/71	6.43 (4.86)	2.34 (2.14)
Cyst >3 cm	50/71	6.77 (4.90)	2.47 (2.23)
Enhanced cyst wall	8/48	8.78 (6.12)	2.01 (0.82)
Main duct 5-9 mm	10/70	8.74 (6.19)	2.35 (1.44)
Nonenhancing mural nodule	1/46	2.03	14.3
Abrupt change in caliber	4/71	6.20 (4.13)	2.40 (1.28)
High risk stigmata	26/71	5.73 (4.66)	2.55 (2.70)
Jaundice	7/69	7.75 (5.57)	3.21 (1.67)
Enhancing mural nodule	15/71	3.90 (3.61)	2.64 (3.34)
Main duct >10 mm	13/70	5.03 (4.35)	1.81 (1.10)
No worrisome features or high risk stigmata	7/71	8.93 (7.02)	2.95 (1.67)

Chapter 4. The lysosomal aminopeptidase tripeptidyl peptidase 1 displays increased activity in malignant pancreatic cysts

4.1 Abstract

Incidental detection of pancreatic cysts has increased dramatically over the last decade, but risk stratification and clinical management remain a challenge. Mucinous cysts are precursor lesions to pancreatic cancer, however, the majority are indolent. Current diagnostics cannot identify mucinous cysts that harbor cancer or reliably differentiate these lesions from nonmucinous cysts, which present minimal risk of malignant progression. We previously determined that activity of two aspartyl proteases, gastricsin and cathepsin E, was highly increased in mucinous cysts. Using a global protease activity profiling technology, termed multiplex substrate profiling by mass spectrometry (MSP-MS), we now show that aminopeptidase activity is also elevated in mucinous cysts. Proteomic analysis identified the lysosomal serine protease, tripeptidyl peptidase 1 (TPP1), in cyst fluid and parallel reaction monitoring-based mass spectrometry demonstrated that this protease was significantly more abundant in mucinous cysts. In a cohort of 110 cyst fluid samples, TPP1 activity was increased more than 3-fold in mucinous cysts relative to nonmucinous cysts. Moreover, TPP1 activity is primarily associated with mucinous cysts that harbor high-grade dysplasia or invasive carcinoma. Measurement of TPP1 activity may improve early detection and treatment of these high-risk pancreatic cysts.

4.2 Introduction

Pancreatic cysts are incidentally detected in 13%-45% of patients being evaluated by abdominal MRI [1,2]. Clinical management is confounded by unreliable risk stratification of the malignant

potential of pancreatic cysts. Mucinous cysts, which include intraductal papillary mucinous neoplasms (IPMNs) and mucinous cystic neoplasms (MCNs), are precursor lesions to pancreatic cancer and should be resected if they harbor high-grade dysplasia or invasive cancer (HGD/IC). However, a significant portion of mucinous cysts only contain low-grade dysplasia (LGD). These lesions are generally considered benign and it is currently recommended that they be followed through surveillance. Serous cystadenomas (SCAs) and pancreatic pseudocysts, which are both types of nonmucinous cysts, are also common and present minimal risk to patients if they remain asymptomatic. Unfortunately, it remains challenging to preoperatively diagnose the cyst type or stage of malignant progression to determine if surgical intervention is warranted.

Current management guidelines for pancreatic cysts are largely based on clinical and radiographic features [3]. These guidelines demonstrate variable sensitivity and unsatisfactory specificity for diagnosing cysts with HGD/IC [4]. In an effort to improve diagnostic accuracy, cyst fluid is now routinely collected by endoscopic ultrasound with fine needle aspiration (EUS-FNA) and subject to analysis. Evaluation of the tumor marker, carcinoembryonic antigen (CEA), is 60%-75% sensitive and 84%-93% specific for differentiating nonmucinous from mucinous cysts [5–7]. However, CEA levels are unable to distinguish mucinous cysts with LGD from those with HGD/IC. Cytological assessment of cyst fluid collected by EUS-FNA is also commonly performed. Although it is highly specific for identifying mucinous cysts with HGD/IC, it suffers from a sensitivity of only 30%-50% [8,9].

The limitations of clinicoradiographic and the cyst fluid diagnostics have spurred significant interest in identifying novel molecular biomarkers. Several cyst fluid biomarkers have shown promise for differentiating mucinous from nonmucinous cysts and for determining which mucinous cysts harbor HGD/IC. Examples of these include microRNA [10,11], mucins [12,13],

glucose levels [14,15], DNA methylation [16], telomerase activity [17], and an array of DNA mutations [18–21]. In a recent study, we used a global protease activity profiling technology to identify two aspartyl proteases, gastricsin and cathepsin E, as promising biomarkers for differentiating mucinous from nonmucinous cysts [5]. Analysis of gastricsin activity using a simple, fluorescence-based assay was 95% accurate for classifying mucinous cysts. Immunohistochemical analysis revealed that gastricsin expression in mucinous cysts was primarily associated with regions of LGD; however, activity levels were unable to differentiate LGD from HGD/IC.

In the present study, we identify increased aminopeptidase activity in fluid from mucinous cysts and determine that the lysosomal protease TPP1 is primarily responsible for this activity. Using both a highly sensitive, targeted proteomics workflow and a simple, fluorescence-based assay, we demonstrate that TPP1 levels are significantly increased in mucinous cysts relative to nonmucinous cysts. Interestingly, TPP1 activity is primarily associated with mucinous cysts with HGD/IC, which is a critical factor when determining if a cyst should be surgically resected.

4.2 Materials and methods

Sample acquisition

The present study included 110 cyst fluid samples from patients seen at the University of California San Francisco (San Francisco, CA), the University of Pittsburgh Medical Center (Pittsburgh, PA), Indiana University School of Medicine (Indianapolis, Indiana), and Stanford University School of Medicine (Stanford, CA). All patients were preconsented under

institutional review board approved protocols. Only samples from patients that underwent surgical resection and pathological examination of their cystic lesion were included. Patient information is summarized in Table S1. Information includes the cyst type, highest degree of dysplasia, method of collection, and institution. All samples were stored at -80 °C prior to analysis and subject to a maximum of two freeze-thaw cycles.

Multiplex substrate profiling by mass spectrometry assay

The MSP-MS assay was performed as described previously [5,22]. Cyst fluid protein concentration was first determined through the BCA assay (Thermo Fisher, 23225). Cyst fluid samples were then diluted to 100 μg/mL in pH 3.5 acetate buffer or pH 7.5 phosphate buffer. The 228 tetradecapeptide was split into two pools of 114 peptides each and diluted to 1 μmol/L in either acetate or phosphate buffer. Equal volumes of diluted cyst fluid and peptide pools were combined and incubated at room temperature. After 15 and 60 minutes, 30 μL aliquots were removed and protease activity was quenched with 8 mol/L guanidinium hydrochloride. Aliquots were then desalted using C18 tips (Rainin). The following inhibitors were included in specified MSP-MS assays: 1 mmol/L AEBSF (Sigma, A8456), 2 mmol/L E-64 (Sigma, E3132), 2 μmol/L pepstatin (Sigma, P5318), 2 mmol/L 1,10-phenanthroline (Sigma, 131337), and 10 μmol/L AAF-CMK (Enzo, BML-PI123).

Mass spectrometry analysis was performed with an LTQ Orbitrap XL Mass Spectrometer (Thermo Fisher) coupled to an Ultra Performance Liquid Chromatography (UPLC) System (Waters). Peptides were separated over a C18 column (Thermo Fisher, ES800) with a 65-minute linear gradient from 2%-30% acetonitrile. MS spectra were acquired over an m/z range of 325-1,500 and MS/MS spectra were obtained for the six most intense precursor ions by collision-

induced dissociation (CID). Peak lists were generated using MSConvert and searched in Protein Prospector v. 5.10.0 against a database containing the sequences from the 228 tetradecapeptide library. Searches used a mass tolerance of 20 ppm for precursors and 0.8 Da for fragments. The following variable modifications were allowed: N- terminal pyroglutamate conversion from glutamine or glutamate and oxidation of tryptophan, proline, and tyrosine. Search outputs were then processed using the MSP-xtractor software (http://www.craiklab.ucsf.edu/extractor.html). This software extracts the P4-P4' sequences and spectral counts for all identified cleavages. Spectral counts were used for relative quantification of MSP-MS results.

Proteomic analysis of cyst fluid samples

Prior to proteomic analysis, abundant serum proteins were depleted from cyst fluid samples to improve detection of low abundance proteases. First, 5 μL of each cyst fluid sample was added to columns containing resin slurry for immunodepletion of the top 12 most abundant serum proteins (Thermo Fisher, 85164). Samples were incubated with the resin for 1 hour at room temperature. Columns were then placed in collection tubes and centrifuged for two minutes at 1000 g to collect unbound proteins. Protein concentration of the eluate was then determined by BCA assay (Thermo Fisher, 23225).

Serum depleted cyst fluid samples were then processed for proteomic analysis using a standard protocol. Briefly, 5 μg of cyst fluid protein was denatured in 6 mol/L urea, disulfide bonds were reduced with 10 mmol/L DTT, and free thiols were then alkylated with 12.5 mmol/L iodoacetamide. The urea concentration was diluted to 2 mol/L using 25 mmol/L ammonium bicarbonate and 100 ng of trypsin was added for 16 hours at 37 °C. Following trypsinization, the sample was desalted using a C18 tip (Rainin), dried, and resuspended in 0.1% formic acid. One

fifth of the cyst fluid sample was then injected onto an Orbitrap Fusion Lumos Mass Spectrometer (Thermo Fisher) coupled to an UPLC System (Waters). A 240 minute linear gradient from 2%-30% acetonitrile was used for peptide separation with a flow rate of 300 nL/min. Survey scans were recorded over a 375-1500 m/z range and the 20 most intense precursor ions from each survey scan were fragmented by high-energy collision dissociation (HCD).

Peak lists were generated from MS/MS data using an in-house software called PAVA and searched in Protein Prospector v. 5.10.0. Peak lists were searched against all human protein sequences in the SwissProt database (downloaded November 1, 2017). This database was concatenated with a fully randomized set of entries to estimate the false discovery rate (FDR). For database searches, peptides sequences were matched to tryptic peptides with up to two missed cleavages. Carbamidomethylation of cysteine residues was used as a constant modification and variable modifications included oxidation of methionine, N-terminal pyroglutamate from glutamine, N-terminal acetylation, and loss of N-terminal methionine. The mass accuracy tolerance was set to 20 ppm for precursor ions and 30 ppm for fragment ions. An FDR of less than 1% was used for all searches.

Parallel reaction monitoring of cyst fluid proteases

To select peptides for parallel reaction monitoring (PRM) assays, we initially performed shotgun proteomic analysis of recombinant TPP1, gastricsin, and cathepsin E. First, 10 ng of recombinant protease was denatured, reduced, alkylated, and digested. Mass spectrometry data was then collected using the same system, method, and search parameters as described above for analysis of cyst fluid proteins. From the shotgun proteomic results, we selected two of the

identified peptides for TPP1, gastricsin, and cathepsin E for inclusion in PRM assays (Table S4). Peptides were prioritized based on whether they were from the mature forms of the proteases and if they had been previously used in targeted proteomics assays [23].

For PRM assays, 5 µL of cyst fluid protein was serum depleted as described above. The UPLC System was also operated using the same parameters. The MS acquisition method consisted of a full MS1 scan event followed by six targeted MS/MS scans for the peptides from TPP1, gastricsin, and cathepsin E. A 0.8 Da mass window was used for precursor ion isolation. The MS1 scan was performed at a resolving power of 120,000 while the MS/MS scans were performed at a resolving power of 30,000.

Relative quantitation of gastricsin and cathepsin E peptides was performed using the Skyline software package. Quantitation was based on the area under the curve of the eight most intense transitions for each peptide. To correct for potential differences in protein loading between runs, peak areas were normalized by the median peak area of all fragmented ions from shotgun proteomic analysis carried out using the same sample. The average area under the curve of the two peptides from each protease was then used to estimate the protein abundance in a given cyst fluid sample.

Peptide synthesis

Solid-phase synthesis of internally quenched fluorescent substrates was carried out using standard Fmoc chemistry on a Syro II automated peptide synthesizer (Biotage). The peptides contain a C-terminal Fmoc-Lys(dinitrophenol) (Anaspec, A23856) and a Lys-(7-methoxcoumarin-4-acetic acid)-OH (EMD Millipore, 852095) in either the P3, P2, or P1 position. Amino acids were coupled to Wang resin preloaded with the Fmoc-Lys(dinitrophenol).

The following conditions were used for coupling of Fmoc protected amino acids: 5 equivalents of amino acid, 4.9 equivalents of HCTU, and 20 equivalents of N-methylmorpholine. Peptides were cleaved from the resin by treatment with a solution of 95% trifluoroacetic acid, 2.5% Triisopropylsilane, and 2.5% water for 1 hour. Peptides were precipitated in cold diethyl ether and dried. Crude peptides were purified by HPLC and the chemical composition was confirmed through LC-MS analysis.

TPP1 activity analysis using internally quenched fluorescent substrates

Fluorescence-based TPP1 activity assays were performed in triplicate in black, round-bottom 384-well plates. A final volume of 15 μL of pH 4.5 acetate buffer was used for all assays. The pH of 4.5 was selected on the basis of promoting TPP1 activity [24]. For analysis of TPP1 activity in cyst fluid samples, 2 μmol/L of pepstatin was included in the acetate buffer to inhibit residual aspartyl protease activity. Substrate concentration was 20 μmol/L unless otherwise stated. Recombinant TPP1 activity was assessed at 2 nmol/L, except for when determining the limit of detection when concentrations down to 1.5 pmol/L were assessed. Cyst fluid assays were carried out using 0.75 μL of sample per well. Substrate cleavage was monitored over 1 hour with a Synergy HT Plate Reader (Biotek) using excitation and emission wavelengths of 328nm and 393nm, respectively. Activity of cyst fluid samples and recombinant TPP1 is expressed as the initial velocity of substrate hydrolysis in relative fluorescent units per second (RFU/sec).

Statistical testing and data analysis

Two-tailed *t*-tests and two-way ANOVAs were used for assessing differences in protease activity between cysts. The specific test is indicated in text. All mass spectrometry data was log2

transformed prior to statistical testing. Logistic regression models were employed for cyst prediction and generating receiver operating characteristic (ROC) curves. Youden's J statistic was used for identifying the optimal activity cutoff for assessing sensitivity and specificity of markers. RStudio was used to generate heatmaps, volcano plots, Venn diagrams, and ROC curves. GraphPad Prism was used for scatter plots, bar charts, and to fit kinetic data. Protease substrate specificity was visualized using iceLogo software [25]. Gene Ontology (http://geneontology.org/) was used for annotating proteases from proteomics data.

4.4 Results

Aminopeptidase activity is enhanced in mucinous cysts

We previously used our MSP-MS assay to assess global proteolytic activity in fluid from 23 pancreatic cysts [5]. In the present study, we analyzed an additional 12 mucinous cysts using our MSP-MS assay. We performed the assay under both acidic conditions and at neutral pH. In line with our previous results, mucinous cysts displayed increased proteolytic activity under acidic conditions with an increase in the average number of detected peptide cleavages relative to nonmucinous cysts (Fig. S1A). In contrast, both cyst types cleaved similar numbers of peptides at neutral pH. We then analyzed which amide bond along the 14-mers from the 228 member peptide library used in MSP-MS was cleaved most frequently. We did not observe any significant differences at neutral pH (Fig. S1B). However, under acidic conditions, a number of amide bonds were cleaved to a greater extent in mucinous cysts and we were particularly interested to observe that cleavage after the third amino acid from the N-terminus was the most significantly enriched (Fig. 1A). This triamino peptidase activity was not inhibited by pepstatin

(Fig. 1B and Fig. S1C), indicating that it was not driven by the pepstatin-sensitive aspartyl proteases identified previously. Pepstatin did decrease the number of detected peptide cleavages at the other positions that were enriched in mucinous cysts at acidic pH (B6, B8, B9, B10). Treatment with several other broad-spectrum protease inhibitors also failed to impact the observed triamino peptidase activity (Fig. S1C), however, a tripeptide chloromethylketone inhibitor almost completely ablated this (Fig. 1B). Halomethylketone-based inhibitors are generally used to target cysteine and serine proteases, providing further evidence that non-aspartyl, protease activity is increased in mucinous cysts [26].

We also analyzed all 93 detected triamino peptidase cleavage events to determine if individual triamino peptidase cleavages were enriched in mucinous cysts (Fig. 1C). Indeed, nine unique events met our selectivity criteria for differentiating mucinous from nonmucinous cysts (+/- $\log_2$(mucinous/nonmucinous), $P < 0.05$). Collectively, these results suggest that non-aspartyl, triamino peptidase activity is increased in mucinous cysts and specific peptide cleavages might serve as a biomarker.

Proteomic analysis identifies increased levels of TPP1 in mucinous pancreatic cysts

We next sought to identify the specific protease responsible for the increased triamino peptidase activity. In our previous study, we performed shotgun proteomic analysis of several cyst fluid samples and identified three aminopeptidases. However, all were metallopeptidases and, based on our inhibitor sensitivity data (Fig. S1C), are likely not driving the observed triamino peptidase activity. To increase the likelihood of identifying the target aminopeptidase here, we first performed a depletion step to remove the abundant serum proteins that are commonly detected in pancreatic cyst fluid. We then performed proteomic analysis and detected a total of 1,632

proteins between six cyst fluid samples (Table S2). Gene ontology analysis revealed that 139 of these proteins are proteases, including 19 aminopeptidases (Table S3). Fifty-three of the identified proteases were only detected in mucinous cysts, while nine were selectively present in nonmucinous cysts (Fig. 2A). As expected, gastricsin and cathepsin E were only identified in mucinous cysts. The majority of aminopeptidases were also only present in mucinous cysts and none were exclusively found in nonmucinous cysts (Fig. 2B and Table S3).

To determine which of the pepstatin-insensitive aminopeptidases was responsible for the increased triamino peptidase activity we considered the following features: catalytic class, pH optimum, and which type of cyst they were detected in (Table S3). Only seven of the identified aminopeptidases were serine or cysteine proteases, which could be targeted by the halomethylketone-based inhibitor (Fig. 1B). Of these, only TPP1 and TPP2 are reported to be inhibited by the specific tripeptide chloromethylketone that we used [27,28]. However, TPP2 is most active at neutral pH, whereas TPP1, which is normally localized to the lysosome, is known to be active under the acidic conditions where we observed increased triamino peptidase activity [24]. TPP1 was also only detected in fluid from mucinous cysts, providing further evidence that it is driving the enhanced activity.

In order to confirm TPP1's increased abundance in mucinous cysts, we developed a PRM assay for relative quantitation of this protease through mass spectrometry. PRM enables the highly sensitive, analysis of targeted peptides for relative quantitation of proteins of interest. We also developed PRM assays for gastricsin and cathepsin E in order to compare the relative fold change in abundance for all the proteases we identified with increased activity in mucinous cysts. Our final PRM assay targeted two peptides per protease for a total of six peptides (Table S4). As expected based on our activity and shotgun proteomics data, TPP1 displayed significantly

increased abundance in mucinous cysts (Fig. 2C). The $\log_2$ (fold change) in abundance for TPP1 was 8.2, while for gastricsin and cathepsin E it was 11.6 and 9.5, respectively (Fig. S2).

TPP1 activity is increased in mucinous cysts

Due to the increased abundance of TPP1 observed through targeted proteomic analysis, we decided to further pursue this protease as a putative biomarker for pancreatic cysts. To this end, we wanted to develop a simple, fluorescent assay for activity analysis in cyst fluid. Several fluorescent substrates have been reported for TPP1, however, we decided to leverage our MSP-MS assay to identify novel substrate sequences that might have improved turnover rates and improve the sensitivity for protease detection in clinical samples [24,29]. MSP-MS analysis of recombinant TPP1 confirmed that this protease readily accommodates hydrophobic amino acids in the P1 and P1' positions, which flank the cleavage site (Fig. 3A). The P1 position also showed a slight preference for glutamine and aspartic acid. We then selected three TPP1 substrates from our MSP-MS library and incorporated the P3 – P4' sequences into internally quenched fluorescent peptides. These substrates were selected based on several features. First, we identified substrates that showed time-dependent cleavage by recombinant TPP1 in our MSP-MS assay (Fig. S3A). Substrates also needed to generally follow the P1 and P1' specificity of TPP1. We then prioritized substrates whose cleavage differentiated mucinous from nonmucinous cysts in our initial low pH MSP-MS analysis of cyst fluid and are likely indicative of TPP1 activity (Fig. 3B and Fig. S3B).

For synthesis of internally quenched substrates containing the selected P3 – P4' sequences, we wanted to maintain the free N-terminal amine for recognition by TPP1. Therefore, the quencher was appended to the C-terminus while the fluorophore was conjugated to the side-

chain amine of a lysine residue. We then tested which position on the nonprime-side of the scissile bond best accommodated this lysine-fluorophore (Fig. 3C and Fig. S3C). For our most rapidly cleaved substrate, this was the P1 position where the bulky lysine-fluorophore replaced a tryptophan residue. This substrate demonstrated a kcat/Km of 2.5 $\mu mol^{-1}s^{-1}$ and we could use this to quantitate as little as 95 pg/mL of TPP1 (Fig. S4).

We next used this substrate to directly assess the levels of TPP1 activity in 110 cyst fluid samples, including the 35 that were previously analyzed by MSP-MS. Cleavage of our TPP1 substrate was increased approximately 3-fold in mucinous cysts relative to nonmucinous cysts (Fig. 4A). However, 27 mucinous cysts cleaved the TPP1 substrate at a lower rate than what was on average observed in nonmucinous cysts. Analysis of the receiver operating characteristic (ROC) curve for TPP1 activity revealed an area under the curve (AUC) of 0.72. At the optimal cutoff of 1.1 RFU/sec, the sensitivity and specificity for TPP1 activity analysis was 62% and 80%, respectively. This performance is comparable to what is generally reported for CEA [4], although within our cohort CEA at the standard cutoff of 192 ng/mL demonstrated a sensitivity of 65% and specificity of 94% [5]. Addition of TPP1 activity analysis to our previously identified aspartyl protease biomarkers did not lead to significantly improved diagnostic performance (Table S5).

We also analyzed whether TPP1 activity differentiated the different types of mucinous and nonmucinous cysts (Fig. S5). SCNs generally had the lowest levels of TPP1 activity, although this was not significant relative to pseudocysts, which where were the other type of nonmucinous cyst included in this study. MCNs tended to have the highest TPP1 activity, however, this was not significant relative to IPMNs.

We next examined whether TPP1 activity levels were associated with the degree of dysplasia within a mucinous cyst. As mentioned previously, this is a critical distinction, as only cysts with HGD/IC generally need to be resected. Mucinous cysts with HGD/IC had a nearly 2-fold increase in TPP1 activity relative to cysts with LGD (Fig. 4C). The ROC curve for TPP1 activity displayed an AUC of 0.65 with a sensitivity of 89% and specificity of 40% (Fig. 4D). This performance is comparable, if not superior, to the ability of consensus guidelines to identify malignant mucinous cysts [30–32].

4.5 Discussion

More than 80% of patients with newly diagnosed pancreatic cancer present with advanced disease, where surgical removal is no longer an option [33]. Early detection of pancreatic cysts presents a unique opportunity for curative resection of this highly lethal disease. Unfortunately, current management guidelines and diagnostic tools are unable to definitively differentiate the cysts that are most likely to progress to pancreatic cancer. Up to two thirds of resected pancreatic cysts are benign or only contain LGD with little risk of malignant progression, thus exposing these patients to unnecessary risk of surgical morbidity and mortality [4,31,34–36].

In a previous study, we identified gastricsin and cathepsin E as promising, cyst fluid-based biomarkers for differentiating mucinous from nonmucinous pancreatic cysts [5]. Here, we applied MSP-MS to search for additional proteases that had both increased activity in mucinous cysts and could differentiate cysts based on their degree of dysplasia. We determined that activity of the aminopeptidase TPP1 is significantly increased in mucinous cysts, but does not differentiate these lesions as well as gastricsin, cathepsin E, or CEA. Within mucinous cysts, TPP1 activity is most highly elevated when HGD/IC is present. As a stand-alone marker, TPP1

activity had modest 89% sensitivity and 40% specificity for differentiating mucinous cysts with HGD/IC from those with LGD. However, this diagnostic performance compares favorably to the clinical and radiologic features, which are commonly assessed as part of current management guidelines [4]. TPP1 may improve the diagnostic performance of other promising cyst fluid biomarkers that are emerging when used in combination [16,18,21,37]. This will be a primary focus of future work investing TPP1 activity in pancreatic cysts.

TPP1 is best studied in the context of the childhood neurodegenerative disease, classic late-infantile form of neuronal ceroid lipofuscinoses (CLN2) and has only rarely been associated with tumorigenesis [38]. CLN2 is a lysosomal storage disease that is driven by mutations in TPP1 that lead to reduced protease activity [39]. Like many lysosomal hydrolases, TPP1 is most active under acidic conditions and its triamino peptidase activity has a pH optimum of 4.5 [24]. In line with our previous paper, our results here continue to point to increased abundance of acid-activated, lysosomal proteases in mucinous pancreatic cysts. This observation may be related to the increased reliance on lysosomal function for nutrient scavenging that is observed with pancreatic intraepithelial neoplasia-derived pancreatic cancer [40–42]. Many acid-activated, lysosomal proteases exhibit increased expression in these tumors and are secreted into the surrounding microenvironment. The abundance of lysosomal proteases in fluid from mucinous cysts suggests that these enzymes are also important for maintenance and growth in cyst-derived pancreatic cancer. In further support of this, several other lysosomal proteases were detected through proteomic analysis of the cyst fluid samples included in this study. Additional studies will seek to determine if analysis of these acid-activated proteases improves diagnostic performance when combined with our current markers.

In conclusion, we demonstrate here that acid-activated aminopeptidase activity is elevated in mucinous pancreatic cysts and this is driven by the lysosomal protease TPP1. Activity analysis of TPP1 is a promising biomarker for differentiating nonmucinous from mucinous cysts and may help address the most critical clinical challenge of identifying mucinous cysts with HGD/IC. Validation of these results has the potential to assist clinical decision making for pancreatic cysts to help ensure appropriate treatment of these challenging precursor lesions of pancreatic cancer.

Acknowledgements

4.6 References

[1]   M. Moris, M.D. Bridges, R.A. Pooley, M. Raimondo, T.A. Woodward, Association Between Advances in High - Resolution Cross - Section Imaging Technologies and Increase in Prevalence of Pancreatic Cysts From 2005 to 2014, Clin. Gastroenterol. Hepatol. 14 (2016) 585–593.

[2]   K.S. Lee, A. Sekhar, N.M. Rofsky, I. Pedrosa, Prevalence of incidental pancreatic cysts in the adult population on MR imaging., Am. J. Gastroenterol. 105 (2010) 2079–84.

doi:10.1038/ajg.2010.122.

[3]   M. Tanaka, C. Fernández-del Castillo, V. Adsay, S. Chari, M. Falconi, J.-Y. Jang, W. Kimura, P. Levy, M.B. Pitman, C.M. Schmidt, M. Shimizu, C.L. Wolfgang, K. Yamaguchi, K. Yamao, International consensus guidelines 2012 for the management of IPMN and MCN of the pancreas., Pancreatology. 12 (2012) 183–97. doi:10.1016/j.pan.2012.04.004.

[4]   J.M. Scheiman, J.H. Hwang, P. Moayyedi, American gastroenterological association technical review on the diagnosis and management of asymptomatic neoplastic pancreatic cysts, Gastroenterology. 148 (2015) 824–48.e22. doi:10.1053/j.gastro.2015.01.014.

[5]   S.L. Ivry, J.M. Sharib, D.A. Dominguez, N. Roy, S.E. Hatcher, M. Yip-Schneider, C.M. Schmidt, R.E. Brand, W.G. Park, M. Hebrok, G. Kim, A.J. O'Donoghue, K.S. Kirkwood, C.S. Craik, Global protease activity profiling provides differential diagnosis of pancreatic cysts., Clin. Cancer Res. (2017). doi:10.1158/1078-0432.CCR-16-2987.

[6]   W.R. Brugge, K. Lewandrowski, E. Lee-Lewandrowski, B.A. Centeno, T. Szydlo, S. Regan, C.F. Del Castillo, A.L. Warshaw, Diagnosis of Pancreatic Cystic Neoplasms: A Report of the Cooperative Pancreatic Cyst Study, Gastroenterology. 126 (2004) 1330–1336. doi:10.1053/j.gastro.2004.02.013.

[7]   W.G. Park, R. Mascarenhas, M. Palaez-Luna, T.C. Smyrk, D.O. Kane, D. Ph, J.E. Clain, J. Michael, R.K. Pearson, B.T. Petersen, M.D. Topazian, S.S. Vege, S.T. Chari, Diagnostic Performance Of Cyst Fluid Carcinoembryonic Antigen And Amylase In Histologically Confirmed Pancreatic Cysts, Pancreas. 40 (2011) 42–45.

doi:10.1097/MPA.0b013e3181f69f36.DIAGNOSTIC.

[8]   A. V. Maker, L.S. Lee, C.P. Raut, T.E. Clancy, R.S. Swanson, Cytology from pancreatic cysts has marginal utility in surgical decision-making, Ann. Surg. Oncol. 15 (2008) 3187–3192. doi:10.1245/s10434-008-0110-0.

[9]   L.A. Van Der Waaij, H.M. Van Dullemen, R.J. Porte, Cyst fluid analysis in the differential diagnosis of pancreatic cystic lesions : a pooled analysis, Gastrointest. Endosc. 62 (2005) 383–389.

[10]   H. Matthaei, D. Wylie, M.B. Lloyd, M.D. Molin, J. Kemppainen, S.C. Mayo, C.L. Wolfgang, R.D. Schulick, L. Langfield, B.F. Andruss, A.T. Adai, R.H. Hruban, A.E. Szafranska-Schwarzbach, A. Maitra, miRNA biomarkers in cyst fluid augment the diagnosis and management of pancreatic cysts, Clin. Cancer Res. 18 (2012) 4713–4724. doi:10.1158/1078-0432.CCR-12-0035.

[11]   J. Wang, P.L. Paris, J. Chen, V. Ngo, H. Yao, M.L. Frazier, A.M. Killary, C.-G. Liu, H. Liang, C. Mathy, S. Bondada, K. Kirkwood, S. Sen, Next generation sequencing of pancreatic cyst fluid microRNAs from low grade-benign and high grade-invasive lesions., Cancer Lett. 356 (2015) 404–9. doi:10.1016/j.canlet.2014.09.029.

[12]   Z. Cao, K. Maupin, B. Curnutte, B. Fallon, C.L. Feasley, E. Brouhard, R. Kwon, C.M. West, J. Cunningham, R. Brand, P. Castelli, S. Crippa, Z. Feng, P. Allen, D.M. Simeone, B.B. Haab, Specific glycoforms of MUC5AC and endorepellin accurately distinguish mucinous from nonmucinous pancreatic cysts., Mol. Cell. Proteomics. 12 (2013) 2724–34. doi:10.1074/mcp.M113.030700.

[13]   J. Sinha, Z. Cao, J. Dai, H. Tang, K. Partyka, G. Hostetter, D.M. Simeone, Z. Feng, P.J. Allen, R.E. Brand, B.B. Haab, A Gastric Glycoform of MUC5AC Is a

Biomarker of Mucinous Cysts of the Pancreas, PLoS One. 11 (2016) e0167070. doi:10.1371/journal.pone.0167070.

[14]   R.A. Carr, M.T. Yip-Schneider, R.E. Simpson, S. Dolejs, J.G. Schneider, H. Wu, E.P. Ceppa, W. Park, C.M. Schmidt, Pancreatic cyst fluid glucose: Rapid, inexpensive, and accurate diagnosis of mucinous pancreatic cysts, Surgery. 163 (2017) 600–605. doi:10.1016/j.surg.2017.09.051.

[15]   T. Zikos, K. Pham, R. Bowen, A.M. Chen, S. Banerjee, S. Friedland, M.M. Dua, J. a Norton, G. a Poultsides, B.C. Visser, W.G. Park, Cyst Fluid Glucose is Rapidly Feasible and Accurate in Diagnosing Mucinous Pancreatic Cysts., Am. J. Gastroenterol. 110 (2015) 909–14. doi:10.1038/ajg.2015.148.

[16]   T. Hata, M. Dal Molin, S.-M. Hong, K. Tamura, M. Suenaga, J. Yu, H. Sedogawa, M.J. Weiss, C.L. Wolfgang, A.M. Lennon, R.H. Hruban, M.G. Goggins, Predicting the grade of dysplasia of pancreatic cystic neoplasms using cyst fluid DNA methylation markers, Clin. Cancer Res. (2017) clincanres.2244.2016. doi:10.1158/1078-0432.CCR-16-2244.

[17]   T. Hata, M. Dal Molin, M. Suenaga, J. Yu, M. Pittman, M.J. Weiss, M. Canto, C.L. Wolfgang, A.M. Lennon, R.H. Hruban, M.G. Goggins, Cyst fluid telomerase activity predicts the histologic grade of cystic neoplasms of the pancreas, Clin. Cancer Res. (2016) 5141–5151. doi:10.1158/1078-0432.CCR-16-0311.

[18]   S. Springer, Y. Wang, M. Dal Molin, D.L. Masica, Y. Jiao, I. Kinde, A. Blackford, S.P. Raman, C.L. Wolfgang, T. Tomita, N. Niknafs, C. Douville, J. Ptak, L. Dobbyn, P.J. Allen, D.S. Klimstra, M.A. Schattner, C.M. Schmidt, M. Yip-Schneider, O.W. Cummings, R.E. Brand, H.J. Zeh, A.D. Singhi, A. Scarpa, R. Salvia, G. Malleo, G.

Zamboni, M. Falconi, J.Y. Jang, S.W. Kim, W. Kwon, S.M. Hong, K.B. Song, S.C. Kim, N. Swan, J. Murphy, J. Geoghegan, W. Brugge, C. Fernandez-Del Castillo, M. Mino-Kenudson, R. Schulick, B.H. Edil, V. Adsay, J. Paulino, J. Van Hooft, S. Yachida, S. Nara, N. Hiraoka, K. Yamao, S. Hijioka, S. Van Der Merwe, M. Goggins, M.I. Canto, N. Ahuja, K. Hirose, M. Makary, M.J. Weiss, J. Cameron, M. Pittman, J.R. Eshleman, L.A. Diaz, N. Papadopoulos, K.W. Kinzler, R. Karchin, R.H. Hruban, B. Vogelstein, A.M. Lennon, A Combination of Molecular Markers and Clinical Features Improve the Classification of Pancreatic Cysts, Gastroenterology. 149 (2015) 1501–1510. doi:10.1053/j.gastro.2015.07.041.

[19]   J. Wu, H. Matthaei, A. Maitra, M. Dal Molin, L.D. Wood, J.R. Eshleman, M. Goggins, M.I. Canto, R.D. Schulick, B.H. Edil, C.L. Wolfgang, A.P. Klein, L.A. Diaz, P.J. Allen, C.M. Schmidt, K.W. Kinzler, N. Papadopoulos, R.H. Hruban, B. Vogelstein, Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development., Sci. Transl. Med. 3 (2011) 92ra66. doi:10.1126/scitranslmed.3002543.

[20]   J. Wu, Y. Jiao, M. Dal Molin, A. Maitra, R.F. de Wilde, L.D. Wood, J.R. Eshleman, M.G. Goggins, C.L. Wolfgang, M.I. Canto, R.D. Schulick, B.H. Edil, M. a Choti, V. Adsay, D.S. Klimstra, G.J. a Offerhaus, A.P. Klein, L. Kopelovich, H. Carter, R. Karchin, P.J. Allen, C.M. Schmidt, Y. Naito, L. a Diaz, K.W. Kinzler, N. Papadopoulos, R.H. Hruban, B. Vogelstein, Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways., Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 21188–93. doi:10.1073/pnas.1118046108.

[21]   A.D. Singhi, K. McGrath, R.E. Brand, A. Khalid, H.J. Zeh, J.S. Chennat, K.E. Fasanella, G.I. Papachristou, A. Slivka, D.L. Bartlett, A.K. Dasyam, M. Hogg, K.K. Lee,

J.W. Marsh, S.E. Monaco, N.P. Ohori, J.F. Pingpank, A. Tsung, A.H. Zureikat, A.I. Wald, M.N. Nikiforova, Preoperative next-generation sequencing of pancreatic cyst fluid is highly accurate in cyst classification and detection of advanced neoplasia, Gut. (2017) gutjnl-2016-313586. doi:10.1136/gutjnl-2016-313586.

[22]   A.J. O'Donoghue, A.A. Eroy-reveles, G.M. Knudsen, J. Ingram, M. Zhou, J.B. Statnekov, A.L. Greninger, D.R. Hostetter, G. Qu, D.A. Maltby, M.O. Anderson, J.L. Derisi, J.H. Mckerrow, A.L. Burlingame, C.S. Craik, Global identification of peptidase specificity by multiplex substrate profiling, Nat. Methods. 9 (2012) 1095–1100. doi:10.1038/NMETH.2182.

[23]   U. Kusebauch, D.S. Campbell, E.W. Deutsch, C.S. Chu, D.A. Spicer, M.Y. Brusniak, J. Slagel, Z. Sun, J. Stevens, B. Grimes, D. Shteynberg, M.R. Hoopmann, P. Blattmann, A. V. Ratushny, O. Rinner, P. Picotti, C. Carapito, C.Y. Huang, M. Kapousouz, H. Lam, T. Tran, E. Demir, J.D. Aitchison, C. Sander, L. Hood, R. Aebersold, R.L. Moritz, Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome, Cell. 166 (2016) 766–778. doi:10.1016/j.cell.2016.06.041.

[24]   Y. Tian, I. Sohar, J.W. Taylor, P. Lobel, Determination of the substrate specificity of tripeptidyl-peptidase I using combinatorial peptide libraries and development of improved fluorogenic substrates, J. Biol. Chem. 281 (2006) 6559–6572. doi:10.1074/jbc.M507336200.

[25]   N. Colaert, K. Helsens, L. Martens, J. Vandekerckhove, K. Gevaert, Improved visualization of protein consensus sequences by iceLogo., Nat. Methods. 6 (2009) 786–7. doi:10.1038/nmeth1109-786.

[26]   L.E. Sanman, M. Bogyo, Activity-based profiling of proteases., Annu. Rev.

Biochem. 83 (2014) 249–73. doi:10.1146/annurev-biochem-060713-035352.

[27]   L. Lin, I. Sohar, H. Lackland, P. Lobel, The human CLN2 protein/tripeptidyl-peptidase I is a serine protease that autoactivates at acidic pH, J. Biol. Chem. 276 (2001) 2249–2255. doi:10.1074/jbc.M008562200.

[28]   R. Gavioli, T. Frisan, S. Vertuani, G.W. Bornkamm, M.G. Masucci, C-myc overexpression activates alternative pathways for intracellular proteolysis in lymphoma cells, Nat. Cell Biol. 3 (2001) 283–288. doi:10.1038/35060076.

[29]   M.Y. Kondo, I.E. Gouvea, D.N. Okamoto, J.A.N. Santos, C. Souccar, K. Oda, L. Juliano, M.A. Juliano, Analysis of catalytic properties of tripeptidyl peptidase i (TTP-I), a serine carboxyl lysosomal protease, and its detection in tissue extracts using selective FRET peptide substrate, Peptides. 76 (2016) 80–86. doi:10.1016/j.peptides.2016.01.009.

[30]   P. Kaimakliotis, B. Riff, K. Pourmand, V. Chandrasekhara, E.E. Furth, E.S. Siegelman, J. Drebin, C.M. Vollmer, M.L. Kochman, G.G. Ginsberg, N.A. Ahmad, Sendai and Fukuoka Consensus Guidelines Identify Advanced Neoplasia in Patients With Suspected Mucinous Cystic Neoplasms of the Pancreas, Clin. Gastroenterol. Hepatol. 13 (2015) 1808–1815. doi:10.1016/j.cgh.2015.03.017.

[31]   C.Y. Hsiao, C.Y. Yang, J.M. Wu, T.C. Kuo, Y.W. Tien, Utility of the 2006 Sendai and 2012 Fukuoka guidelines for the management of intraductal papillary mucinous neoplasm of the pancreas: A single-center experience with 138 surgically treated patients, Med. (United States). 95 (2016). doi:10.1097/MD.0000000000004922.

[32]   B.K.P. Goh, D.M.Y. Tan, M.M.F. Ho, T.K.H. Lim, A.Y.F. Chung, L.L.P.J. Ooi, Utility of the Sendai Consensus Guidelines for Branch-Duct Intraductal Papillary Mucinous Neoplasms: A Systematic Review, J. Gastrointest. Surg. 18 (2014) 1350–1357.

doi:10.1007/s11605-014-2510-8.

[33] L. Rahib, J.M. Fleshman, L.M. Matrisian, J.D. Berlin, Evaluation of Pancreatic Cancer Clinical Trials and Benchmarks for Clinically Meaningful Future Trials, JAMA Oncol. 2 (2016) 1209. doi:10.1001/jamaoncol.2016.0585.

[34] C. Correa-Gallego, C.R. Ferrone, S.P. Thayer, J.A. Wargo, A.L. Warshaw, C. Fernandez-Del Castillo, Incidental pancreatic cysts: Do we really know what we are watching?, Pancreatology. 10 (2010) 144–150. doi:10.1159/000243733.

[35] C.E. Parra-herran, M.T. Garcia, L. Herrera, P.A. Bejarano, Cystic Lesions of the Pancreas : Clinical and Pathologic Review of Cases in a Five Year Period, J. Pancreas. 11 (2010) 358–364.

[36] N.P. Valsangkar, V. Morales-Oyarvide, S.P. Thayer, C.R. Ferrone, J.A. Wargo, A.L. Warshaw, C. Fern??ndez-Del Castillo, 851 resected cystic tumors of the pancreas: A 33-year experience at the Massachusetts General Hospital, Surg. (United States). 152 (2012) 4–12. doi:10.1016/j.surg.2012.05.033.

[37] T. Hata, M. Dal Molin, M. Goggins, Cyst fluid telomerase activity predicts the histologic grade of cystic neoplasms of the pancreas, Clin. Cancer Res. (2016) 2003.

[38] A.A. Golabek, E. Kida, Tripeptidyl-peptidase I in health and disease, Biol. Chem. 387 (2006) 1091–1099. doi:10.1515/BC.2006.135.

[39] D.E. Sleat, R.J. Donnelly, H. Lackland, C. Liu, I. Sohar, R.K. Pullarkat, P. Lobel, Association of Mutations in a Lysosomal Protein with Classical Late-Infantile Neuronal Ceroid Lipofuscinosis, Science (80-.). 277 (1997) 1802–1806.

[40] R.M. Perera, N. Bardeesy, Pancreatic cancer metabolism: Breaking it down to build it back up, Cancer Discov. 5 (2015) 1247–1261. doi:10.1158/2159-8290.CD-15-

0671.

[41]   G. Bai, C. Wu, Y. Gao, G. Shu, Exploring the Functional Disorder and Corresponding Key Transcription Factors in Intraductal Papillary Mucinous Neoplasms Progression, Int. J. Genomics. 2015 (2015).

[42]   R.M. Perera, S. Stoykova, B.N. Nicolay, K.N. Ross, J. Fitamant, M. Boukhali, J. Lengrand, V. Deshpande, M.K. Selig, C.R. Ferrone, J. Settleman, G. Stephanopoulos, N.J. Dyson, R. Zoncu, S. Ramaswamy, W. Haas, N. Bardeesy, Transcriptional control of autophagy-lysosome function drives pancreatic cancer metabolism., Nature. (2015). doi:10.1038/nature14587.
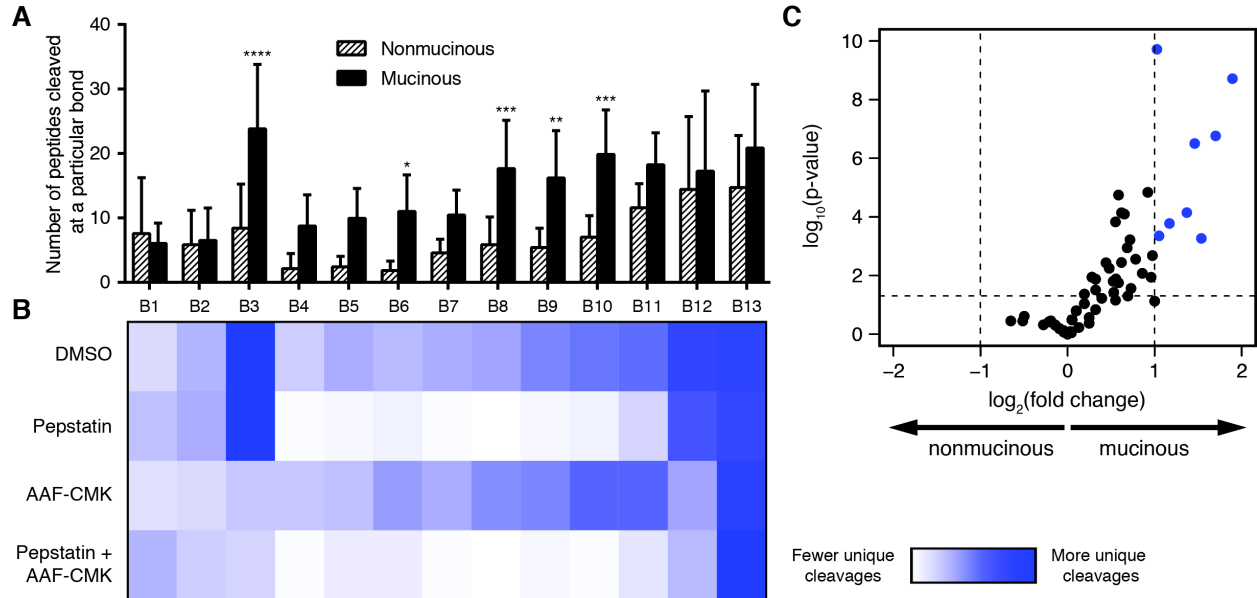
Figure 4.1 Identification of low pH aminopeptidase activity in mucinous cysts through MSP-MS. (**A**) The average number of detected cleavages between each amide bond (BX) within the 228 peptide library when the MSP-MS assay was performed at pH 3.5. A representative 14-mer library peptide is shown below the bar chart with black circles indicating amino acids. Two-way ANOVA was used for determining statistical significance (*$P < 0.05$, **$P < 0.005$, ***$P < 0.0005$, ****$P < 0.00005$) (**B**) Heatmap displaying the number of cleavages at each amide bond following treatment of a mucinous cyst fluid sample with DMSO, pepstatin, Ala-Ala-Phe-chloromethylketone (AAF-CMK), or a combination. (**C**) Volcano plot comparing the amount of cleavage at the third amide bond for individual library peptides. We detected cleavage of 93 library peptides at this bond. Spectral counts of cleavage products were used for quantification of the fold change (mucinous/nonmucinous) and hypothesis testing. Dotted lines indicate $P < 0.05$ and ± 1 log$_2$(fold change) and blue dots are used for triamino peptidase cleavages that met these criteria.
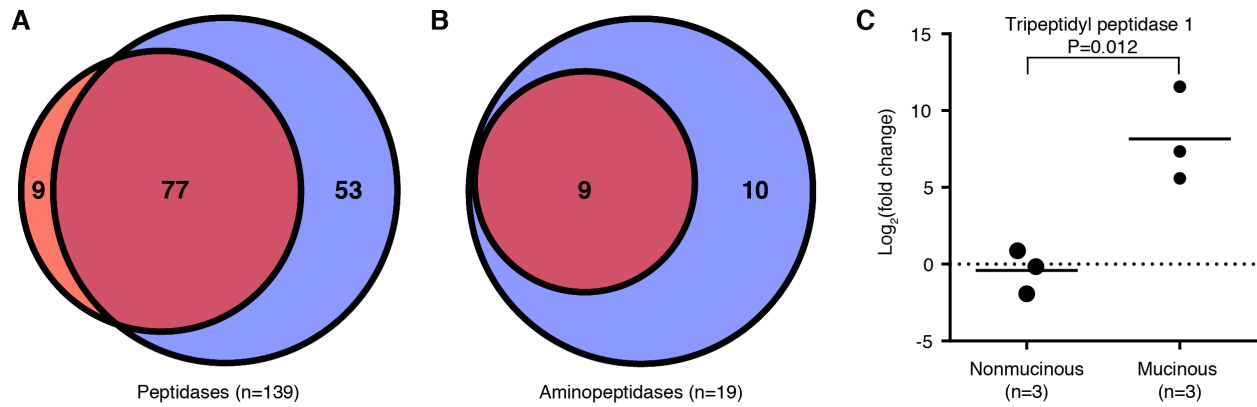
Figure 4.2 TPP1 abundance is increased in fluid from mucinous cysts. Venn diagrams showing the peptidases (**A**) and aminopeptidases (**B**) identified through shotgun proteomic analysis of fluid from mucinous (blue, n=3) and nonmucinous (red, n=3) cysts. (**C**) PRM analysis TPP1 peptides from nonmucinous and mucinous cyst fluid. Integrated peak areas of the eight most abundant transitions were used to determine the fold change (mucinous/nonmucinous). An unpaired, two-tailed *t*-test of the log$_2$(fold change) was used for determining statistical significance.
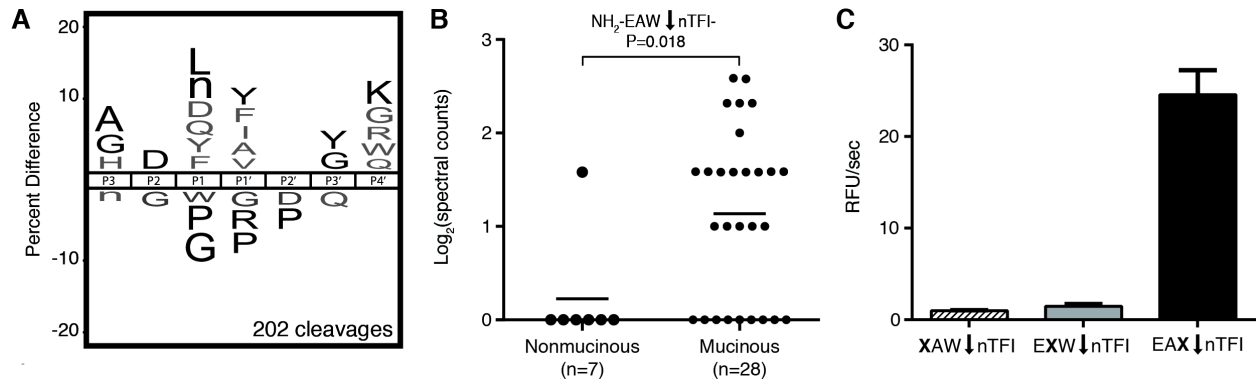
Figure 4.3 Development of an internally quenched fluorescent substrate for TPP1 activity analysis in cyst fluid. (**A**) iceLogo depicting the substrate specificity of recombinant TPP1 based on the 202 peptide cleavages detected through MSP-MS analysis. Amino acids that are significantly enriched at specific positions are shown in black (P < 0.05), while those that trended towards being enriched are in grey (P < 0.20). (**B**) Cleavage of a peptide from the MSP-MS library in nonmucinous and mucinous cysts. Arrow indicates that cleavage is occurring after the third amino acid from the N-terminus and the 7 remaining C-terminal amino acids aren't shown. Spectral counts were used for relative quantification and an unpaired, two-tailed *t*-test was used for statistical testing. (**C**) Cleavage of internally quenched fluorescent substrates by recombinant TPP1. **X** indicates the position of the lysine-fluorophore.
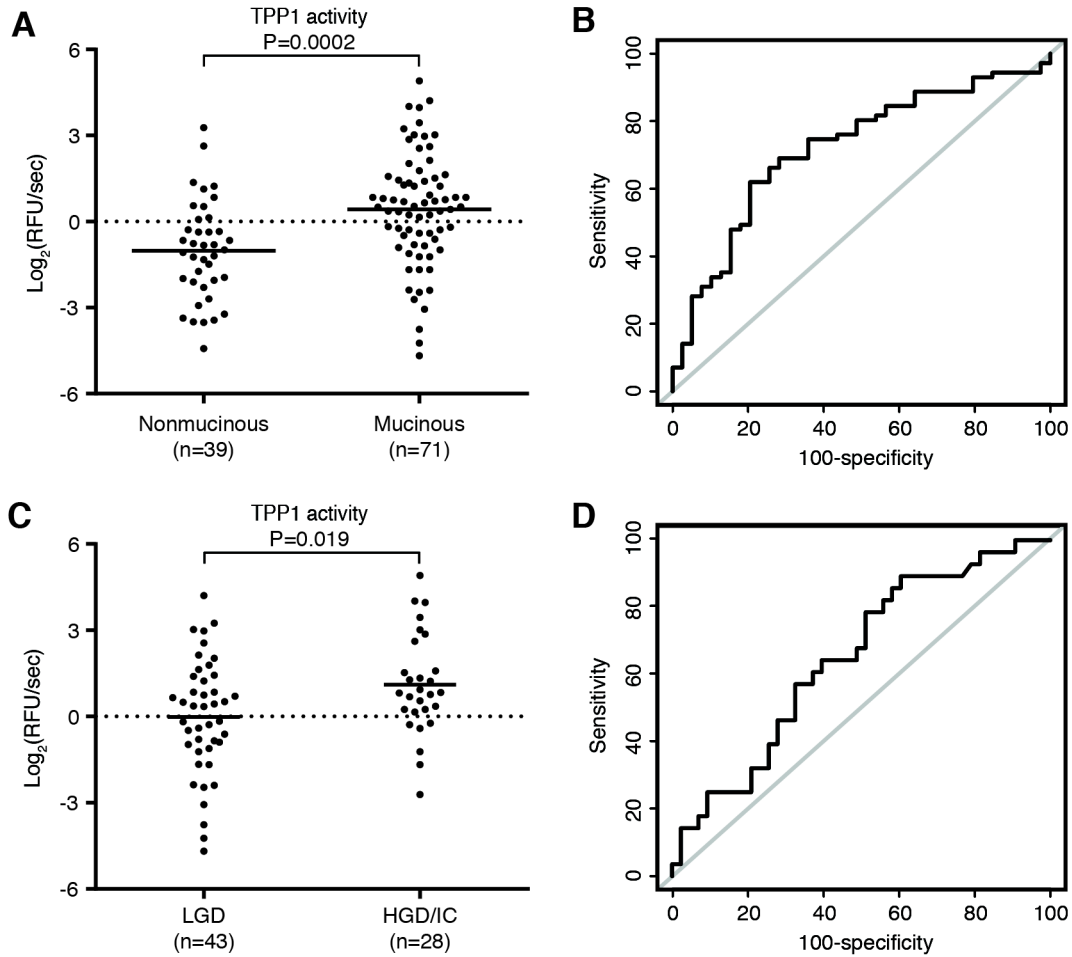
Figure 4.4 Analysis of TPP1 activity in cyst fluid samples. (**A**) Comparison of TPP1 activity levels in nonmucinous and mucinous cysts and ROC curve (**B**) for differentiating these cyst types. (**C**) TPP1 activity in mucinous cysts of varying degrees of dysplasia and associated ROC curve (**D**) for distinguishing LGD from HGD/IC. For all statistical testing, an unpaired, two-tailed *t*-test was used.
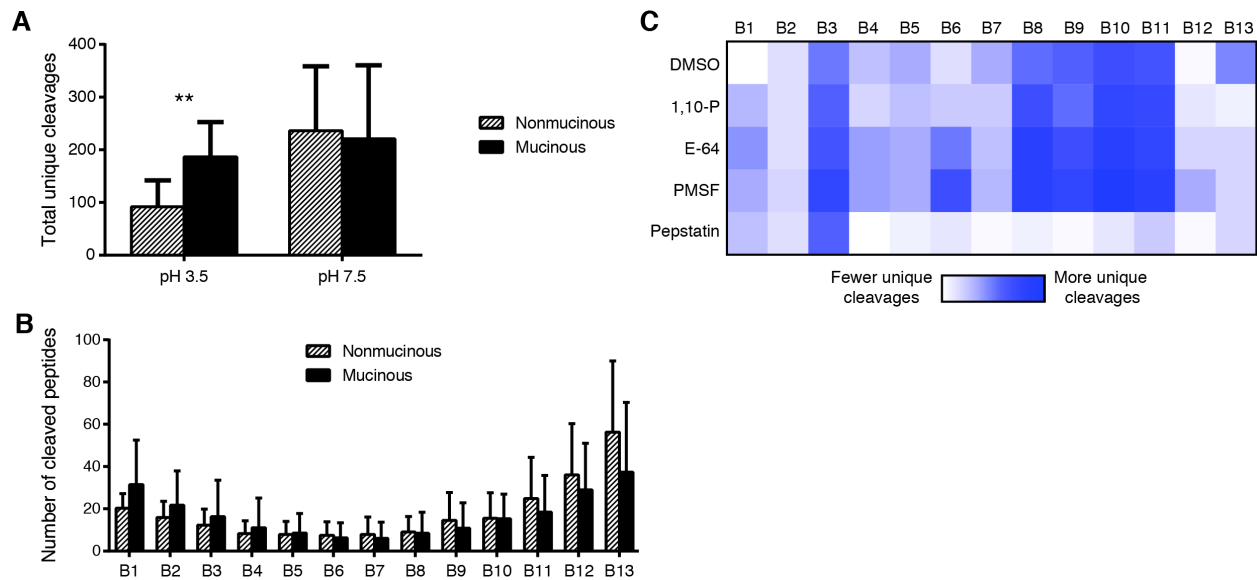
Supplementary Figure 4.1 Non-aspartyl protease activity is enriched in mucinous cysts. (A) Average number of detected peptide cleavages during MSP-MS analysis at pH 3.5 and 7.5. **(B)** The average number of detected cleavages between each amide bond when the MSP-MS assay was performed at pH 7.5. **(C)** Heatmap showing the number of detected cleavages at each amide bond following treatment of a mucinous cyst with DMSO or a panel of broad-spectrum protease inhibitors. For all analysis, two-way ANOVA was used for determining statistical significance (**P < 0.005).

Supplementary Figure 4.2 PRM analysis of peptides from PGC and CTSE in fluid from nonmucinous (NM, n=3) and mucinous (M, n=3) cysts. Peak areas of the eight most abundant transitions were used to determine the fold change and assess statistical significance using unpaired, two-tailed *t*-tests.
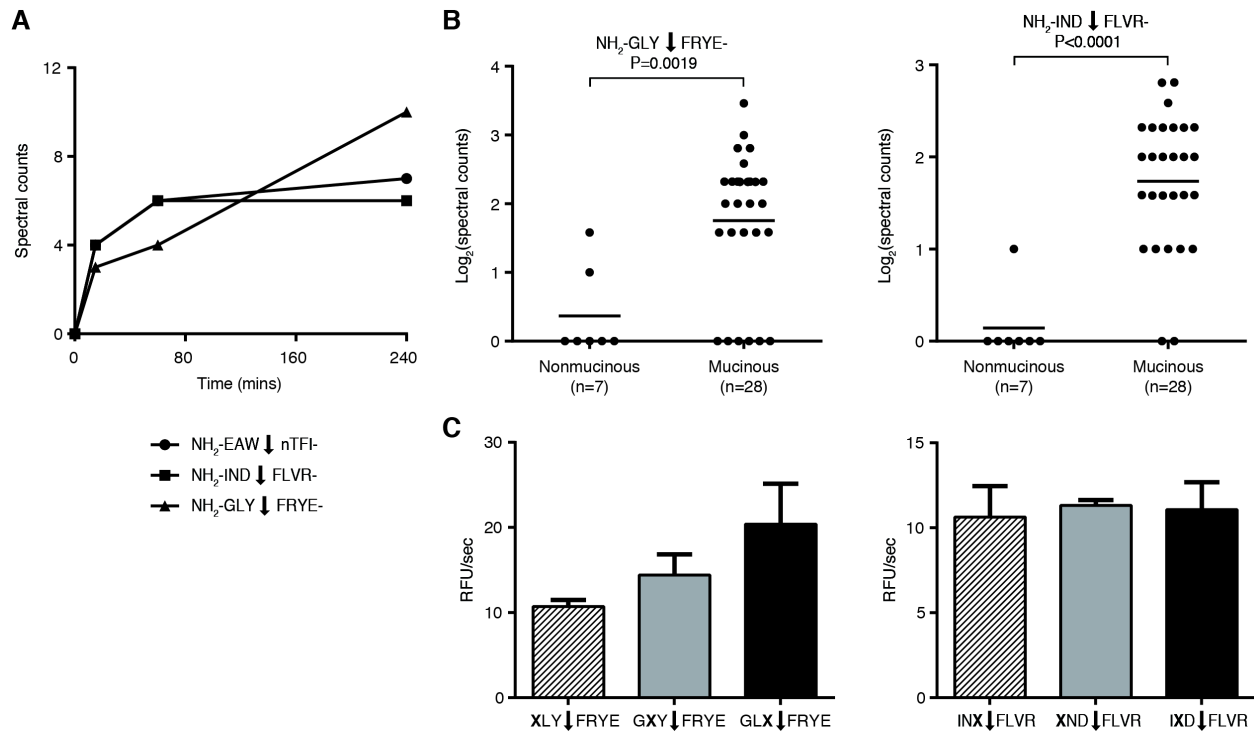
Supplementary Figure 4.3 TPP1 fluorescent substrate design. (A) Time dependent cleavage of 3 peptides in the MSP-MS library. Arrow indicates that cleavage is after the third amino acid and the 7 remaining C-terminal amino acids aren't shown. **(B)** Triamino peptidase cleavage of two peptides from **A** by cyst fluid samples. An unpaired, two-tailed *t*-test was used for statistical testing. **(C)** Cleavage of internally quenched fluorescent substrates by TPP1. The position of the lysine-fluorophore is indicated by **X.**
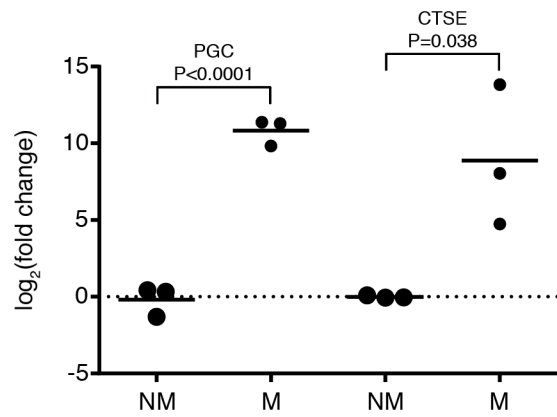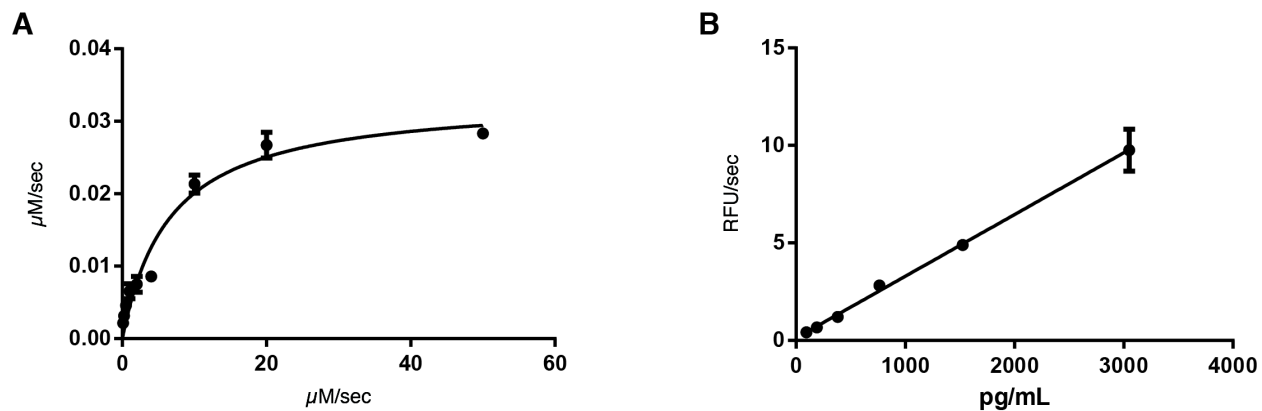
Supplementary Figure 4.4 (**A**) Kinetic analysis of TPP1 cleavage of our most sensitive internally quenched fluorescent substrate. TPP1 concentration was constant at 2 nmol/L, while substrate concentration ranged from 0.1-50 μmol/L. (**B**) Determination of the limit of detection of TPP1 activity using internally quenched fluorescent substrate. Substrate concentration was 20 μmol/L and TPP1 concentration ranged from 95-3000 pg/mL.
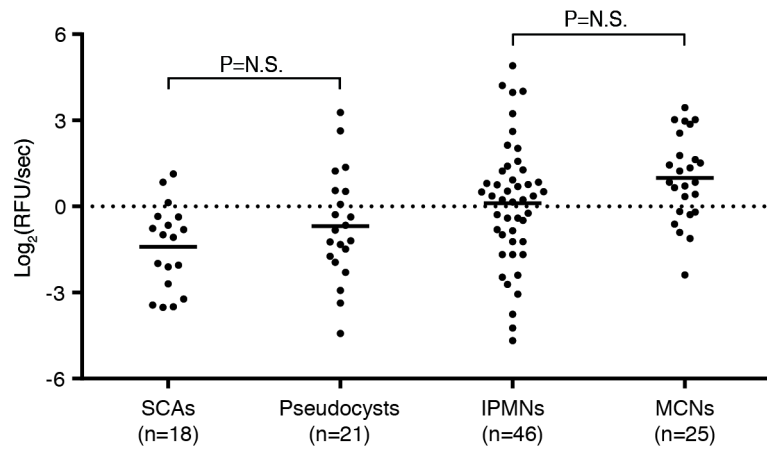
Supplementary Figure 4.5 TPP1 activity in different types of cysts. Statistical testing was done with an unpaired, two-tailed *t*-test.

Tables
Supplementary Table 4.1 Characteristics of 110 patients analyzed in this study.

Internal ID	Diagnosis	Highest grade of dysplasia	Cyst size (mm)	Collection method (EUS/Surgery)	Age	Gender	Institution
14	IPMN	High	22	Surgery	60	Female	UCSF
15	IPMN	High	19	Surgery	69	Male	UCSF
23	IPMN	Low	8	Surgery	55	Female	UCSF
34	IPMN	High	59	Surgery	78	Male	UCSF
55	IPMN	High	26	Surgery	69	Female	Stanford
138	IPMN	Low	65	EUS	68	Male	Stanford
770	IPMN	High	39	Surgery	72	Female	Stanford
775	IPMN	High	120	Surgery	78	Male	Stanford
788	IPMN	High	18	Surgery	64	Male	Stanford
797	IPMN	High	30	Surgery	79	Female	Stanford
801	IPMN	High	25	Surgery	68	Male	Stanford
41	IPMN	Low	18	Surgery	41	Male	Pittsburgh
47	IPMN	Moderate	31	Surgery	47	Female	Pittsburgh
51	IPMN	High	20	Surgery	51	Male	Pittsburgh
52	IPMN	Moderate	47	Surgery	52	Female	Pittsburgh
53	IPMN	Low	25	Surgery	53	Female	Pittsburgh
54	IPMN	Moderate	38	Surgery	54	Female	Pittsburgh
62	IPMN	Moderate	40	Surgery	62	Female	Pittsburgh
64	IPMN	Moderate	47	Surgery	64	Male	Pittsburgh
69	IPMN	Moderate	30	Surgery	69	Female	Pittsburgh
77	IPMN	Moderate	65	Surgery	77	Female	Pittsburgh
85	IPMN	High	27	Surgery	85	Female	Pittsburgh
90	IPMN	High	NA	EUS	69	Female	Indiana
169	IPMN	High	35	ERCP	75	Male	Indiana
1177	IPMN	High	57	Surgery	67	Female	Indiana
1183	IPMN	High	120	Surgery	78	Male	Indiana
1187	IPMN	High	NA	Surgery	66	Female	Indiana
1209	IPMN	High	50	Surgery	74	Male	Indiana
1217	IPMN	High	23	Surgery	58	Male	Indiana
1233	IPMN	High	NA	Surgery	71	Female	Indiana
1250	IPMN	Low	23	Surgery	68	Female	Indiana
1251	IPMN	Low	25	Surgery	79	Male	Indiana
1252	IPMN	High	11	Surgery	72	Male	Indiana
1253	IPMN	High	30	Surgery	63	Female	Indiana
1255	IPMN	Low	37	Surgery	80	Female	Indiana
1272	IPMN	Low	20	Surgery	40	Female	Indiana
1275	IPMN	High	20	Surgery	62	Male	Indiana
157	IPMN	Low	35	Surgery	84	Male	Stanford
685	IPMN	Low	4	Surgery	71	Female	Stanford
733	IPMN	Low	45	EUS	78	Male	Stanford
755	IPMN	Moderate	60	Surgery	60	Male	Stanford
760	IPMN	Moderate	35	Surgery	84	Female	Stanford
761	IPMN	Moderate	30	Surgery	74	Female	Stanford
785	IPMN	Low	90	Surgery	72	Male	Stanford
786	IPMN	Moderate	50	Surgery	71	Female	Stanford
4	MCN	Moderate	23	Surgery	57	Female	UCSF
6	MCN	High	22	Surgery	74	Female	UCSF
125	MCN	High	57	Surgery	54	Female	UCSF
130	MCN	High	125	Surgery	47	Female	UCSF
30	MCN	Moderate	60	Surgery	48	Male	Stanford
136	MCN	Low	100	Surgery	44	Female	Stanford
139	MCN	Low	140	Surgery	22	Female	Stanford

Supplementary Table 4.1 (continued)

Internal ID	Diagnosis	Highest grade of dysplasia	Cyst size (mm)	Collection method (EUS/Surgery)	Age	Gender	Institution
686	MCN	Low	125	Surgery	44	Female	Stanford
716	MCN	Low	30	Surgery	65	Female	Stanford
720	MCN	Low	45	Surgery	49	Female	Stanford
752	MCN	High	152	Surgery	42	Female	Stanford
768	MCN	Low	65	Surgery	25	Female	Stanford
787	MCN	Low	30	Surgery	76	Female	Stanford
799	MCN	Moderate	70	Surgery	31	Female	Stanford
42	MCN	Moderate	100	Surgery	42	Female	Pittsburgh
50	MCN	High	215	Surgery	50	Female	Pittsburgh
59	MCN	Moderate	19	Surgery	59	Female	Pittsburgh
74	MCN	Low	115	Surgery	74	Female	Pittsburgh
75	MCN	Low	30	Surgery	75	Female	Pittsburgh
1152	MCN	Low	18	Surgery	61	Female	Indiana
1182	MCN	Low	50	Surgery	25	Female	Indiana
1191	MCN	Low	25	Surgery	24	Female	Indiana
1261	MCN	Low	40	Surgery	24	Female	Indiana
1263	MCN	High	28	Surgery	44	Male	Indiana
1265	MCN	Low	30	Surgery	50	Female	Indiana
1269	MCN	Low	50	Surgery	57	Female	Indiana
81	Pseudocyst	NA	55	EUS	48	Male	Stanford
98	Pseudocyst	NA	219	Surgery	30	Female	Stanford
101	Pseudocyst	NA	40	EUS	62	Female	Stanford
126	Pseudocyst	NA	90	EUS	57	Male	Stanford
151	Pseudocyst	NA	12	Surgery	55	Female	Stanford
678	Pseudocyst	NA	100	Surgery	56	Female	Stanford
705	Pseudocyst	NA	90	Surgery	60	Male	Stanford
43	Pseudocyst	NA	79	EUS	43	Male	Pittsburgh
49	Pseudocyst	NA	98	EUS	49	Male	UCSF
63	Pseudocyst	NA	91	EUS	63	Male	Pittsburgh
70	Pseudocyst	NA	100	Surgery	70	Male	Pittsburgh
78	Pseudocyst	NA	31	EUS	78	Female	Pittsburgh
11	Pseudocyst	NA	NA	EUS	68	Female	Indiana
50	Pseudocyst	NA	NA	Surgery	59	Female	Indiana
159	Pseudocyst	NA	NA	EUS	67	Male	Indiana
311	Pseudocyst	NA	NA	EUS	60	Male	Indiana
1142	Pseudocyst	NA	NA	Surgery	52	Male	Indiana
1178	Pseudocyst	NA	NA	Surgery	66	Female	Indiana
1180	Pseudocyst	NA	NA	Surgery	42	Female	Indiana
1191	Pseudocyst	NA	NA	Surgery	24	Female	Indiana
1230	Pseudocyst	NA	NA	Surgery	58	Female	Indiana
21	SCA	NA	20	Surgery	58	Male	UCSF
87	SCA	NA	160	Surgery	62	Male	UCSF
68	SCA	NA	50	EUS	59	Male	Stanford
147	SCA	NA	38	EUS	45	Female	Stanford
679	SCA	NA	65	Surgery	70	Female	Stanford
732	SCA	NA	100	Surgery	43	Female	Stanford
748	SCA	NA	60	Surgery	66	Male	Stanford
767	SCA	NA	105	Surgery	47	Male	Stanford
802	SCA	NA	60	Surgery	75	Female	Stanford
10	SCA	NA	40	Surgery	55	Male	Indiana
121	SCA	NA	110	EUS	33	Female	Indiana
128	SCA	NA	25	EUS	55	Female	Indiana
409	SCA	NA	45	Surgery	26	Female	Indiana
852	SCA	NA	25	Surgery	30	Female	Indiana
1053	SCA	NA	12	Surgery	41	Female	Indiana
1218	SCA	NA	82	Surgery	76	Female	Indiana
1234	SCA	NA	93	Surgery	76	Male	Indiana
1283	SCA	NA	30	Surgery	74	Female	Indiana

Supplementary Table 4.4 Analysis of TPP1, PGC, and CTSE relative abundance by PRM.

Protein	Gene	Accession Number	Peptide	m/z	z	Peak Area					
						NM1	NM2	NM3	M1	M2	M3
Cathepsin E	CTSE	P14091	QFYSVFDR	531.2562	2	2606	11146	13246	239940247	204468	1828912
Cathepsin E	CTSE	P14091	SQLSEFWK	512.7585	2	4301	4068	3415	235387329	70887	4299258
Gastricsin	PGC	P20142	SYYSVYDLGNNR	725.8335	2	1554	3576	3630	13785954	11531615	23011535
Gastricsin	PGC	P20142	GLLGEFLR	452.7662	2	9996	25918	6194	28137132	24801435	64446224
Tripeptidyl-peptidase 1	TPP1	O14773	LYQQHGAGLFDVTR	535.6108	3	1391	4264	14085	12396244	817414	176277
Tripeptidyl-peptidase 1	TPP1	O14773	LFGGNFAHQASVAR	492.2565	3	1150	4447	3669	16709112	752909	288863

Supplementary Table 4.5 Diagnostic performance of TPP1 activity analysis in 110 cyst fluid sample.

Markers	AUC	Sensitivity (%)	Specificity (%)
TPP1	0.721	62	79.5
CTSE	0.828	70.4	92.3
PGC	0.979	93	100
TPP1+CTSE	0.849	74.6	87.2
TPP1+PGC	0.981	93	100
CTSE + PGC	0.987	94.4	100
TPP1+CTSE+PGC	0.987	94.4	100

Chapter 5. Future directions

5.1 Validation of a multimodal protease-based diagnostic for pancreatic cysts

Our work has demonstrated that protease activity is a promising new biomarker for differentiating pancreatic cysts based on their likelihood of malignant transformation. We have already shown diagnostic utility for our activity-based biomarkers in a cohort of 110 patient samples and our immediate next step will be validating diagnostic performance using a blinded patient cohort. We are currently working with a consortium of pancreatic cancer centers to assemble a cohort of between 200-400 patient cyst fluid samples. Prior to analysis of our validation cohort, we will need to establish the optimal cutoff for each activity-marker, determine our limit of detection, standardize sample treatment, and perform various other analytical validations steps.

In parallel, we also plan to develop a colorimetric, chemiluminescent, and enzyme-linked immunosorbent assay (ELISA) for each protease target. Colorimetric assays, although not as sensitive as fluorescence, are simple to automate and can be readily adopted into a clinical laboratory setting. ELISAs and chemiluminescence are both highly sensitive and might enable a lower limit of detection, in turn leading to improved diagnostic performance. ELISAs also have the potential for increased selectivity as antibodies tend to be much more specific than peptide substrates for assessment of a single analyte. We will assess each assay format within our training cohort of 110 cyst fluid samples to prioritize which to carry forward to validation. Assay type performance is generally dependent on the marker being targeted and it is possible that multiple formats will be necessary for optimal diagnostic performance.

Our results show that gastricsin and cathepsin E accurately differentiate mucinous from nonmucinous cysts, while TPP1 is better for distinguishing mucinous cysts based on their grade

of dysplasia. These are both critical distinctions that need to be made to ensure appropriate management of pancreatic cysts. Therefore, we anticipate a sequential diagnostic strategy for clinical use of our protease-based biomarkers. First, mucinous cysts will be identified using gastricsin and cathepsin E analysis. TPP1 assessment will then identify those mucinous cysts with HGD/IC that should be resected. As TPP1 alone is only 59% accurate for differentiating mucinous cysts with LGD from those with HGD/IC, we expect that the inclusion of additional biomarkers will significantly improve diagnostic performance. Collectively, these results suggest that a panel of biomarkers, including the protease-based markers identified here, will be required for the appropriate management of pancreatic cysts.

5.2 Protease-activated chemotherapeutics for treatment of pancreatic cancer

Through this work we have identified several proteases whose expression is primarily associated with PDAC and its precursor lesions. We have primarily leveraged the activity of these proteases for the development of diagnostics. We are now beginning to investigate whether PDAC-associated proteases could be used to activate therapeutics within the tumor microenvironment for cancer treatment. More specifically, we plan to synthesize chemotherapeutics that are conjugated to a masking peptide that blocks either cellular permeability or target engagement. The peptide mask is stable in circulation, but can be cleaved by a protease within the tumor microenvironment to release the active drug. Similar strategies have been previously employed for both small molecules and biologics in cancer [1–3]. These efforts have shown promise for improving the therapeutic window of their parent agents. However, most efforts have been limited to targeting the mask to a single tumor-associated protease. Through MSP-MS, we have obtained a more global view of PDAC-associated

proteolysis and believe this will allow us to design masking peptides that are highly selective for multiple proteases within the tumor microenvironment. In combination with the diagnostic work proposed above, we believe these efforts may lead to highly personalized treatment regimens for pancreatic cancer patients.
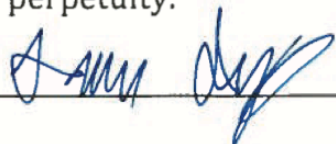
5.3 References

[1]   W. Wu, Y. Luo, C. Sun, Y. Liu, P. Kuo, J. Varga, R. Xiang, R. Reisfeld, K.D. Janda, T.S. Edgington, C. Liu, Targeting cell-impermeable prodrug activation to tumor microenvironment eradicates multiple drug-resistant neoplasms, Cancer Res. 66 (2006) 970–980. doi:10.1158/0008-5472.CAN-05-2591.

[2]   S.R. Denmeade, A.M. Mhaka, D.M. Rosen, W.N. Brennen, S. Dalrymple, I. Dach, C. Olesen, B. Gurel, A.M. Demarzo, G. Wilding, M.A. Carducci, C.A. Dionne, J. V Møller, P. Nissen, S.B. Christensen, J.T. Isaacs, Engineering a Prostate-Specific Membrane Antigen – Activated Tumor Endothelial Cell Prodrug for Cancer Therapy, Sci. Transl. Med. 4 (2012).

[3]   L.R. Desnoyers, O. Vasiljeva, J.H. Richardson, A. Yang, E.E.M. Menendez, T.W. Liang, C. Wong, P.H. Bessette, K. Kamath, S.J. Moore, J.G. Sagert, D.R. Hostetter, F. Han, J. Gee, J. Flandez, K. Markham, M. Nguyen, M. Krimm, K.R. Wong, S. Liu, P.S. Daugherty, J.W. West, H.B. Lowman, Tumor-specific activation of an EGFR-targeting probody enhances therapeutic index, Sci. Transl. Med. 5 (2013). doi:10.1126/scitranslmed.3006682.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature_____ Date __6/19/18__