Characterization of glycoproteins in pancreatic cyst fluid using a high performance multiple lectin affinity chromatography platform

Francisca O. Gbormittah[1], Brian B. Haab[2], Katie Partyka[2], Carolina Garcia-Ott[1], Marina Hancapie[3], William S. Hancock[1*]

[1] Barnett Institute and Department of Chemistry and Chemical Biology, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA

[2] Centers for Cancer Genomics and Quantitative Biology, Van Andel Research Institute, Grand Rapids, Michigan 49503, USA

[3] Genzyme, a Sanofi Company, 45 New York Avenue, Framingham, MA. 01701

Supplementary Figure 1. Workflow diagram showing experimental process used in glycoproteomic studies of two analysis sample set. Pancreatic cyst fluid samples were purified from immunoglobulins (A, D, E, G, M, and light chains) and albumin using an immobilized antibody HPLC packed PEEK column. Glycoprotein enrichment followed by one dimensional gel electrophoresis were used as further fractionation steps leading to nano-LC-MS/MS analysis of M-LAC bound and unbound fractions.

Supplementary Figure 2. 1D SDS-PAGE of two sample sets used for glycoproteomics analysis. Variations in albumin levels and individual differences in each sample set accounts for variability in proteins identified in sample set one compared to sample set two.

Supplementary Figure 3. MS/MS fragmentation of diagnostic peptide TYAYLFSHPSR of CEL-long isoform. Sequence coverage is 13.95, Charge: +2, Monoisotopic m/z: 671.333 Da (accuracy: +0.05 mmu/+0.08 ppm), ionscore: 44

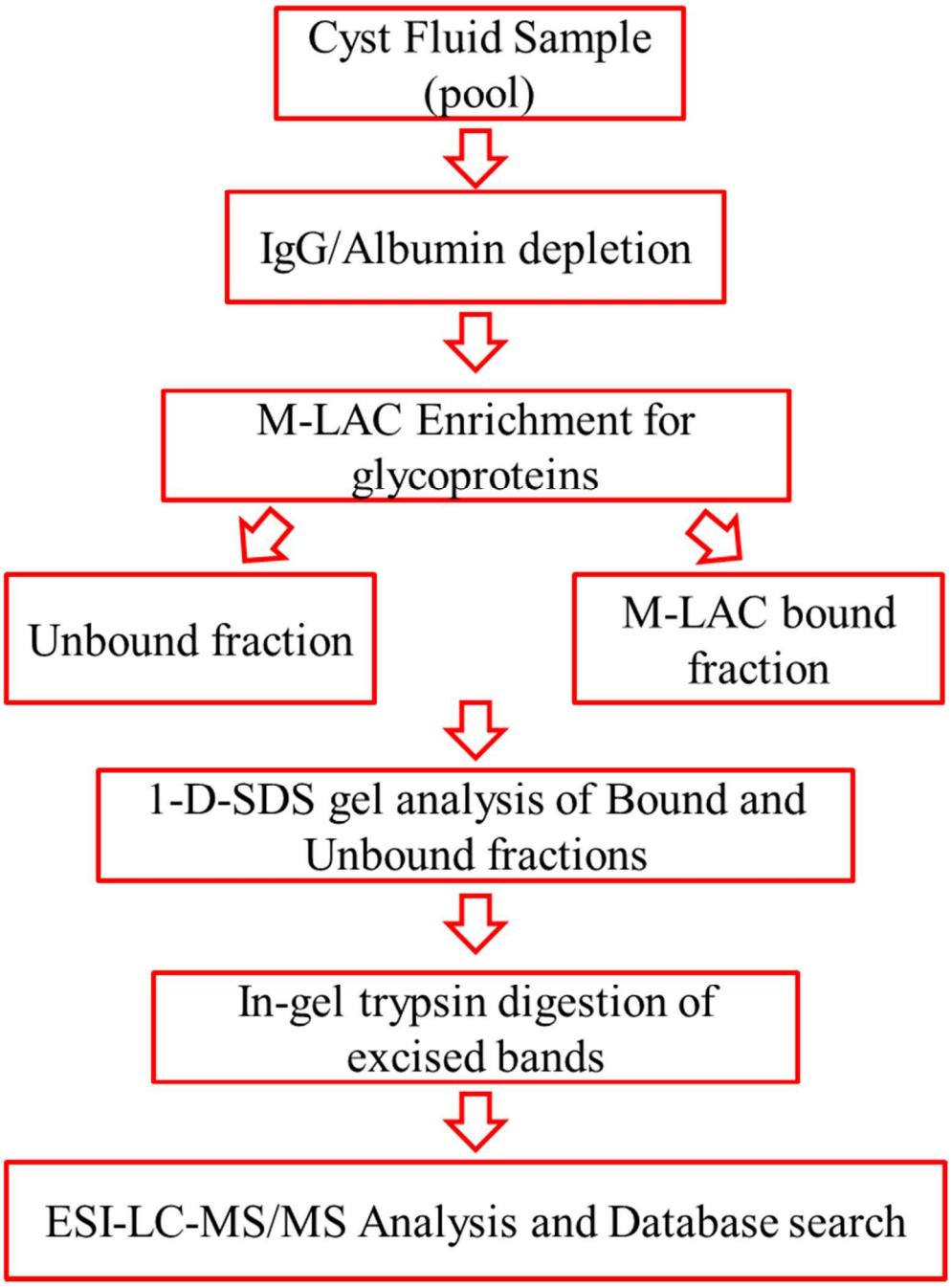
Cyst Fluid Sample (pool)
IgG/Albumin depletion
M-LAC Enrichment for glycoproteins
Unbound fraction
M-LAC bound fraction
1-D-SDS gel analysis of Bound and Unbound fractions
In-gel trypsin digestion of excised bands
ESI-LC-MS/MS Analysis and Database search

Figure 1

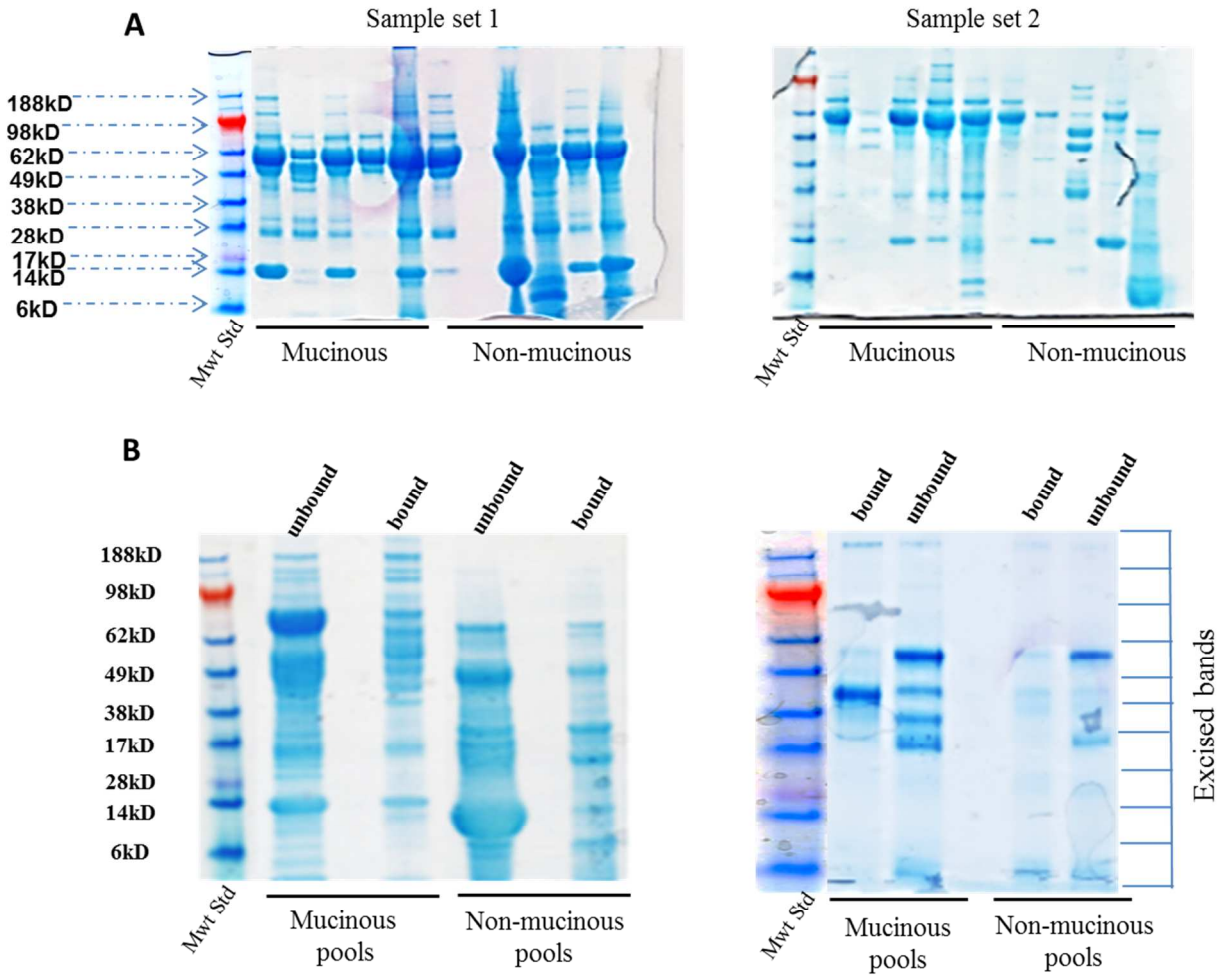
A
Sample set 1
Sample set 2
188kD
98kD
62kD
49kD
38kD
28kD
17kD
14kD
6kD
Mwt Std
Mucinous
Non-mucinous
B
unbound
bound
unbound
bound
bound
unbound
bound
unbound
Mucinous pools
Non-mucinous pools
Excised bands

Figure 2

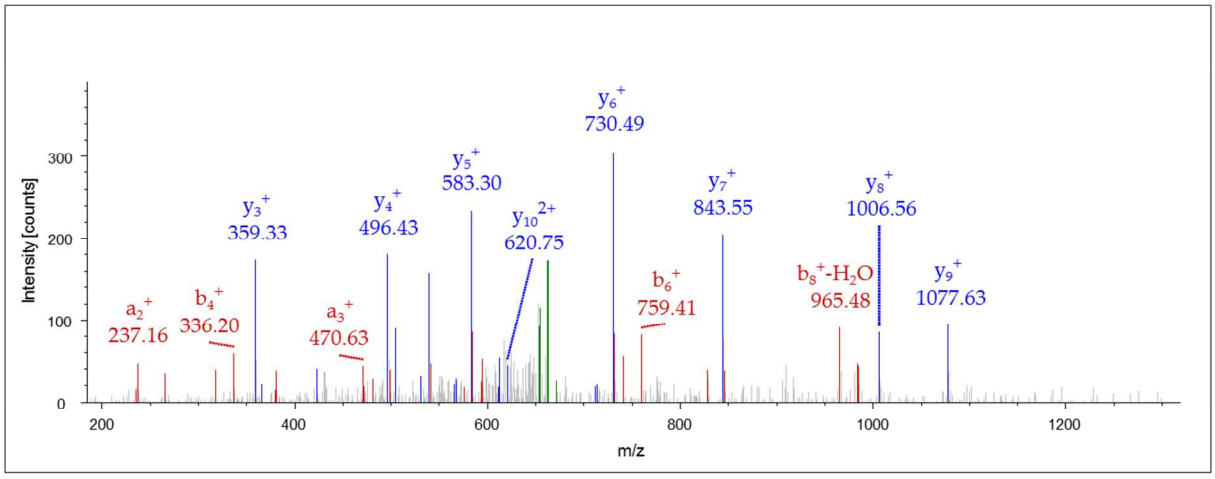
Intensity [counts]
m/z
0
100
200
300
200
400
600
800
1000
1200
a_2^+ 237.16
b_4^+ 336.20
y_3^+ 359.33
a_3^+ 470.63
y_4^+ 496.43
y_5^+ 583.30
y_{10}^{2+} 620.75
y_6^+ 730.49
b_6^+ 759.41
y_7^+ 843.55
b_8^+-H_2O 965.48
y_8^+ 1006.56
y_9^+ 1077.63

Figure 3

Supplementary table **1**. Number of Identified proteins in the Unbound and M-LAC Bound fractions after 1D SDS-PAGE LC-MS/MS Glycoproteomics analysis

Sample set 1	**Number of unbound proteins**			**Number of bound proteins**			**Ratio**
Pools	**Spectral count = 1***	**Spectral count ≥ 2**	**Total proteins**	**Spectral count = 1***	**Spectral count ≥ 2**	**Total proteins**	**(BD/UNBD)X100**
Mucinous	146	223	369	134	160	294	79.7
Non-mucinous	142	290	432	99	184	283	65.5

Sample set 2	**Number of unbound proteins**			**Number of bound proteins**			**Ratio**
Pools	**Spectral count = 1***	**Spectral count ≥ 2**	**Total proteins**	**Spectral count = 1***	**Spectral count ≥ 2**	**Total proteins**	**(BD/UNBD)X100**
Mucinous	93	170	263	80	110	190	72.2
Non-mucinous	71	120	191	44	82	126	66

*Proteins with spectral count equal to one were included in data analysis if the proteins were found with high abundance (spectral count > 10) in one sample set.

Supplementary table **2A**. Identified peptides for bile salt-activated lipase (CEL) long iso-form in M-LAC bound subproteome. Protein sequence coverage is 16.5%.

Sequence	**Charge**	**MH+ [Da]**	**Intensity**	**RT [min]**
ALENPQPHPGWQGTLK	3	1772.909	6.82E+06	44.14
DQHmAIAWVK	3	1214.598	1.59E+06	43.39
GIPFAAPTK	1	901.514	6.70E+06	45.44
KLGLLGDSVDIFK	3	1404.809	4.19E+05	63.46
LGAVYTEGGFVEGVNKK	2	1767.927	1.94E+06	48.20
NPLFWAK	2	875.478	6.99E+06	54.32
TVVDFETDVLFLVPTEIALAQHR	3	2613.397	2.59E+05	89.59
TYAYLFSHPSR	3	1341.660	2.68E+06	48.32
VGCPVGDAAR	2	1001.482	8.43E+06	26.56
VTEEDFYK	2	1030.472	2.22E+06	39.50
AISQSGVALSPWVIQK	2	1683.945	5.78E+05	61.88
LGLLGDSVDIFK	2	1276.716	2.04E+06	70.57
VGPLGFLSTGDANLPGNYGLR	2	2118.103	1.11E+05	70.91

Blue highlight is showing diagnostic peptide sequence unique to bile salt-activated lipase long iso-form.

Supplementary table **2B**. Identified peptides for bile salt-activated lipase (CEL) short iso-form in unbound subproteome. Protein sequence coverage is 13.97%.

Sequence	**Charge**	**MH+ [Da]**	**Intensity**	**RT [min]**
AISQSGVALSPWVIQK	2	1683.943	2.60E+06	60.54
ALTLAYK	1	779.466	2.66E+06	45.25
LGLLGDSVDIFK	2	1276.715	4.94E+06	69.1
NPLFWAK	2	875.477	2.00E+06	55.85
TVVDFETDVLFLVPTEIALAQHR	3	2613.392	7.20E+05	88.57
VGCPVGDAAR	2	1001.483	3.19E+06	31.79
VGPLGFLSTGDANLPGNYGLR	2	2118.099	1.94E+06	70.58

Supplementary table **3**. Novoseek disease relationship to pancreatic cancer and related diseases data of selected protein target list

Protein Name	**Gene Name**	**Disease association**
14-3-3 protein epsilon	YWHAE	-
14-3-3 protein zeta/delta	YWHAZ	-
Adenylyl cyclase-associated protein 1	CAP1	Pancreatic Cancer
Adenylyl cyclase-associated protein 1	CAP1	pancreatitis
Aldo-keto reductase family 1 member B10	AKR1B10	-
Alpha-1-acid glycoprotein 2	ORM2	-
Alpha-2-HS-glycoprotein	AHSG	-
Alpha-amylase 2B	AMY2B	pancreatitis
Annexin A10	ANXA10	-
Annexin A5	ANXA5	pancreatitis
Annexin A5	ANXA5	Pancreatic Ductal Adenocarcinoma
Annexin A5	ANXA5	pancreatic carcinoma
Annexin A5	ANXA5	pancreatic cancer
Aspartate aminotransferase, mitochondrial	GOT2	-
Basement membrane-specific heparan sulfate proteoglycan core protein	HSPG2	-
Bifunctional purine biosynthesis protein PURH	ATIC	-
Bile salt-activated lipase	CEL	pancreas exocrine
Bile salt-activated lipase	CEL	pancreatic tumor
Bile salt-activated lipase	CEL	pancreatitis
Bile salt-activated lipase	CEL	pancreatic cancer
Cadherin-17	CDH17	pancreatitis
Calcium-activated chloride channel regulator 1	CLCA1	pancreatitis
Carbonic anhydrase 1	CA1	chronic pancreatitis
Carbonic anhydrase 2	CA2	chronic pancreatitis
Carboxypeptidase A1	CPA1	Pancreatitis
Carboxypeptidase A2	CPA2	pancreatitis
Carboxypeptidase B	CPB1	pancreatitis alcoholic
Carboxypeptidase B	CPB1	acute pancreatitis
Carboxypeptidase B	CPB1	pancreatitis
Carboxypeptidase B	CPB1	pancreatic cancer
Catalase	CAT	-
Fibronectin	FN1	-
Glutathione S-transferase A1	GSTA1	-
Glutathione S-transferase A2	GSTA2	-

Supplementary table **3** conti.

Protein Name	**Gene Name**	**Disease association**
Glycine amidinotransferase, mitochondrial	GATM	pancreatitis
Heat shock 70 kDa protein 1A/1B	HSPA1A	-
Heat shock 70 kDa protein 1-like	HSPA1L	-
Heat shock 70 kDa protein 6	HSPA6	-
Heat shock cognate 71 kDa protein	HSPA8	-
Hexokinase-1	HK1	Pancreatic Cancer
Hexokinase-1	HK1	pancreatitis
Histone H4	HIST1H4I	-
Interstitial collagenase	MMP1	-
Isoform H14 of Myeloperoxidase	MPO	-
Kininogen-1	KNG1	pancreatitis
Leucine-rich alpha-2-glycoprotein	LRG1	Pancreatitis
Leukotriene A-4 hydrolase	LTA4H	-
Metalloproteinase inhibitor 1	TIMP1	-
Mucin-2	MUC2	pancreatic tumor
Mucin-2	MUC2	pancreatic cancer
Mucin-2	MUC2	carcinoma pancreatic ductal
Mucin-2	MUC2	pancreatic carcinoma
Mucin-2	MUC2	chronic pancreatitis
Mucin-2	MUC2	pancreatic cystadenoma mucinous
Mucin-6	MUC6	pancreatic tumor
Mucin-6	MUC6	chronic pancreatitis
Mucin-6	MUC6	carcinoma pancreatic ductal
Pancreatic alpha-amylase	AMY2A	pancreatitis
Pancreatic lipase-related protein 2	PNLIPRP2	pancreatitis
Pancreatic triacylglycerol lipase	PNLIP	pancreatic insufficiency
Pancreatic triacylglycerol lipase	PNLIP	exocrine pancreatic insufficiency
Pancreatic triacylglycerol lipase	PNLIP	pancreas exocrine
Pancreatic triacylglycerol lipase	PNLIP	pancreatic diseases
Pancreatic triacylglycerol lipase	PNLIP	acute pancreatitis
Pancreatic triacylglycerol lipase	PNLIP	chronic pancreatitis
Pancreatic triacylglycerol lipase	PNLIP	pancreatitis
Pancreatic triacylglycerol lipase	PNLIP	pancreatic cancer

Supplementary table **3** conti.

Protein Name	**Gene Name**	**Disease association**
Periostin	POSTN	Pancreatic Ductal Adenocarcinoma
Periostin	POSTN	Pancreatitis
Periostin	POSTN	Pancreatic cancer
Phosphoglycerate kinase 1	PGK1	-
Pigment epithelium-derived factor	SERPINF1	-
Protein disulfide-isomerase A4	PDIA4	-
Protein S100-A12	S100A12	-
Puromycin-sensitive aminopeptidase	NPEPPS	-
Pyruvate kinase isozymes M1/M2	PKM	pancreatitis
Pyruvate kinase isozymes M1/M2	PKM	pancreatic cancer
Tetranectin	CLEC3B	-
Vinculin	VCL	-
Vitamin D-binding protein	GC	-

(-)'proteins of interest' in which no disease association was found using Novoseek data mining tool.