SUPPLEMENTAL MATERIALS

Methods

Genotyping

Genomic DNA was extracted from either peripheral blood mononuclear cells or from normal tissue obtained from pancreatic resection specimens (duodenum, spleen, or pancreas) as previously described [1]. Sixteen SNPs (in *FUT3, FUT2, FUT6, ABO, GAL3ST2, MUC16, B3GNT3, FAM3B* and *THBS2*) known to be associated with tumor marker levels from prior studies [2-5 6, 7 8] were genotyped (Table S1). The *FUT3* and *FUT2* SNPs used in this study encode for an inactivate protein and the relevant ABO SNP encodes for blood group B. Fourteen SNPs were sequenced using an AmpliSeq Custom Panel (Table S1). Next-generation sequencing (NGS) was performed with 520 chips (Ion S5 system) according to the manufacturer's protocols and as previously described [9, 10]. Genotyping of rs81293 (T202C) in *FUT3* and rs12469459 in *GAL3ST2* was performed with a Taqman™ SNP Genotyping Assay (Applied Biosystems, USA).

ELISA

All assays were performed according to manufacturer's protocols. Absorbance was measured using an xMark™ Microplate Absorbance Spectrophotometer (Bio-Rad Laboratories, Inc. USA). Marker levels were determined from standard curves of the protein standards provided for each ELISA generated from a four-parameter logistic nonlinear regression model using JMP 13 software (SAS Institute, Cary, NC, USA). Thrombospondin-2 was not measured in some Stage I PDAC cases-they were held back from analysis to preserve these samples. Two

different lots of CEA ELISAs were used; the coefficient of variation for lot 1 was 8.9% and for lot 2 was 5.1%. There was a lot-to-lot difference in the mean CEA level in reference serum samples so CEA levels from the 2nd lot were normalized to those of the first lot determined by the mean levels in the reference serum samples in both lots.

Supplemental Results

CA19-9 levels did not differ significantly between *FUT2*-heterozygous and *FUT2*-wild-type subjects. Among *FUT3*-null subjects, CA19-9 levels were not significantly different by FUT2 status. Among FUT2-null subjects CA19-9 levels were not significantly different among *FUT3*-wild-type compared to *FUT3*-heterozygotes.

The overall mean/SD level of CA19-9 in all control patients unclassified with respect to tumor marker genotype in the discovery set was 11.7/11.5 U/ml. We did not find other variants (*FUT6, B3GNT3*)[6] were associated with CA19-9 levels once *FUT3* variant status was accounted for (data not shown).

Among PDAC patients, CA19-9 levels were significantly different within genotype subgroups ($p<0.00001$, Mann-Whitney, FUT2-null versus other groups and FUT3-null versus other groups).

We calculated the number needed to diagnose and the number needed to misdiagnose (NNM)[11]. NNM is an index of diagnostic effectiveness. For CA19-9, the NNM calculated for a PDAC prevalence of 1% is 68, 72 and 75, respectively using the uniform, genotype-stratified, and intact FUT3 genotype-stratified cut-offs. The number needed to diagnose a PDAC using

genotype-stratified CA19-9 alone at a prevalence of 0.5% would be 330; at 1% it would be 164. To detect a Stage I PDAC at a 1% prevalence, it would be 250. Assuming 50% cure rate for patients under regular surveillance diagnosed with PDAC, (based on our CAPS experience[12]and a conservative estimate based on our recent analysis of SEER data in submission), the number needed to test to prevent 1 death would be 500.

In the discovery control set, significant differences in mean CEA levels were observed between the largest group, FUT2+/non-B, vs. each of the other four groups (P=0.0003, 0.0083, 0.0378 and <0.0001, respectively). In the validation set, significant differences in mean CEA levels were again observed between the FUT2+/non-B, and the FUT2-null/B and FUT2+/B groups (each p=0.04)

We tested one candidate *THBS2* variant (rs8089) for its association with serum thrombospondin-2 levels[13], but we did not find evidence for an association (data not shown). The diagnostic sensitivity of genotype-stratified CEA for PDAC was 15% among non-smokers. The diagnostic sensitivity of a uniform cut-off for CA125 for PDAC was 12.8% among non-smokers.

Supplemental References cited

1. Shindo K, Yu J, Suenaga M, et al. Deleterious Germline Mutations in Patients With Apparently Sporadic Pancreatic Adenocarcinoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2017;35(30):3382-90. Epub 2017/08/03.
2. Narimatsu H, Iwasaki H, Nakayama F, et al. Lewis and secretor gene dosages affect CA19-9 and DU-PAN-2 serum levels in normal individuals and colorectal cancer patients. Cancer research. 1998;58(3):512-8. Epub 1998/02/11.
3. Luo G, Guo M, Jin K, et al. Optimize CA19-9 in detecting pancreatic cancer by Lewis and Secretor genotyping. Pancreatology : official journal of the International Association of Pancreatology (IAP) [et al]. 2016;16(6):1057-62. Epub 2016/10/04.
4. Kawai S, Suzuki K, Nishio K, et al. Smoking and serum CA19-9 levels according to Lewis and secretor genotypes. International journal of cancer Journal international du cancer. 2008;123(12):2880-4. Epub 2008/09/23.

5. Park HD, Park KU, Song J, et al. The relationship between Lewis/Secretor genotypes and serum carbohydrate antigen 19-9 levels in a Korean population. The Korean journal of laboratory medicine. 2010;30(1):51-7. Epub 2010/03/04.
6. He M, Wu C, Xu J, et al. A genome wide association study of genetic loci that influence tumour biomarkers cancer antigen 19-9, carcinoembryonic antigen and alpha fetoprotein and their associations with cancer risk. Gut. 2014;63(1):143-51. Epub 2013/01/10.
7. Williams KA, Terry KL, Tworoger SS, et al. Polymorphisms of MUC16 (CA125) and MUC1 (CA15.3) in relation to ovarian cancer risk and survival. PloS one. 2014;9(2):e88334. Epub 2014/02/20.
8. Folkersen L, Fauman E, Sabater-Lleal M, et al. Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. PLoS genetics. 2017;13(4):e1006706. Epub 2017/04/04.
9. Yu J, Sadakari Y, Shindo K, et al. Digital next-generation sequencing identifies low-abundance mutations in pancreatic juice samples collected from the duodenum of patients with pancreatic cancer and intraductal papillary mucinous neoplasms. Gut. 2017;66:1677-87.
10. Suenaga M, Yu J, Shindo K, et al. Pancreatic Juice Mutation Concentrations Can Help Predict the Grade of Dysplasia in Patients Undergoing Pancreatic Surveillance. Clinical cancer research : an official journal of the American Association for Cancer Research. 2018;24(12):2963-74. Epub 2018/01/06.
11. Habibzadeh F, Yadollahie M. Number needed to misdiagnose: a measure of diagnostic test effectiveness. Epidemiology (Cambridge, Mass). 2013;24(1):170. Epub 2012/12/13.
12. Canto MI, Almario JA, Schulick RD, et al. Risk of Neoplastic Progression in Individuals at High Risk for Pancreatic Cancer Undergoing Long-term Surveillance. Gastroenterology. 2018;155(3):740-51.e2. Epub 2018/05/29.
13. Boekholdt SM, Trip MD, Peters RJ, et al. Thrombospondin-2 polymorphism is associated with a reduced risk of premature myocardial infarction. Arteriosclerosis, thrombosis, and vascular biology. 2002;22(12):e24-7. Epub 2002/12/17.

Table S1. SNPs characterized in this study and the associated tumor markers affected

gene	SNP db	position	variant	location	function	genotyping	tumor marker	Chr	Chr_Start	Chr_End
FUT2	rs601338	Chr19:49206674	G428A	exon	nonsense	NGS	CA19-9/CEA	Chr19	49206582	49206721
FUT2	rs281377	Chr19:49206603	T357C	exon	missense	NGS	CA19-9/CEA	Chr19	49206582	49206721
FUT2	rs1047781	Chr19:49206631	A385T	exon	missense	NGS	CA19-9/CEA	Chr19	49206582	49206721
FUT3	rs28362459	Chr19:5844792	T59G	exon	missense	NGS	CA19-9	Chr19	5844736	5844910
FUT3	rs812936	Chr19:5844649	T202C	exon	missense	RT-PCR	CA19-9	Chr19	5844624	5844674
FUT3	rs778986	Chr19:5844537	C314T	exon	missense	NGS	CA19-9	Chr19	5844482	5844648
FUT3	rs3745635	Chr19:5844343	G508A	exon	missense	NGS	CA19-9	Chr19	5844282	5844455
FUT3	rs3894326	Chr19:5843784	T1067A	exon	missense	NGS	CA19-9	Chr19	5843661	5843827
FUT6	rs17271883	Chr19:5834212	g.5834212C>T	intron	missense	NGS	CA19-9	Chr19	5834183	5834331
FUT6	rs3760775	Chr19:5841356	g.5841356G>T	upstream	missense	NGS	CA19-9/CEA	Chr19	5841225	5841380
B3GNT3	rs265548	Chr19:17902334	g.17902334T>C	upstream	missense	NGS	CA19-9	Chr19	17902252	17902426
ABO	rs8176746	Chr9:136131322	c.796C>A	exon	missense	NGS	CEA	Chr9	136131239	136131396
ABO	rs8176719	Chr9:136132908	c.260_262insG	exon	insertion	NGS	CEA	Chr9	136132777	136132946
FAM3B	rs441810	Chr21:42698907	g.42698907A>G	intron	missense	NGS	CEA	Chr21	42698809	42698978
THBS2	rs8089	Chr6:169617726	g.169617726A>C	3'UTR	missense	NGS	thrombospondin-2	Chr6	169617635	169617809
GAL3ST2	rs12469459	Chr2:242716380	c.10A>T	exon	missense	RT-PCR	CA125	Chr2	242716355	242716405

Abbreviations:RT-PCR, real-time PCR

Table S2. SNP stratified tumor marker specificity in validation set controls

CA19-9	Discovery set cut off (Mean + 3SD)	number	false positives (n)	Specificity (95% CI)
FUT3 (-/-)	9.5	22	1	
FUT3 (+/-)	23.5	92	0	
FUT3 (+/+)	38.1	92	0	
FUT2 (-/-)	66.6	53	2	
All genotype stratified				98.8% (96.65% to 99.76%)
Uniform cut-off (not SNP stratified)	46.1	259	4	98.5% (96.09% to 99.58%)

CEA	Discovery set cut off (Mean + 3SD)	number	false positives (n)	Specificity
FUT2+/blood B	6.6	22	0	
FUT2+/non-blood B	2.8	174	3	
FUT2 (-/-)/blood B	4.3	5	0	
FUT2 (-/-)/non blood B	4.0	47	0	
smoker	4.2	11	2	
All genotype stratified				98.1% (95.55% to 99.37%)
Uniform cut-off (Not SNP stratified)	3.9	259	3	98.8% (96.65% to 99.76%)

CA125	Discovery set cut off (Mean + 3SD)	number	false positives (n)	Specificity
GAL3ST2 Wild-type	22.3	81	1	
GAL3ST2 Heterozygous	16.2	85	1	
GAL3ST2 Homozygous	10.7	16	2	
All genotype stratified				97.8% (94.47% to 99.40%)
Uniform cut-off (Not SNP stratified)	19.2	182	4	97.2% (93.71% to 99.10%)

Table S3: Reclassification table of uniform vs. genotype-stratified diagnostic cut-offs

CA19-9	SNP stratified cut-off		
uniform cut-off	negative	positive	Total
Cases			
negative	92	24	116
positive	4	125	129
Total	96	149	245
Controls*			
negative	254	1	255
positive	2	2	4
Total	256	3	259

CA125	SNP stratified cut-off		
uniform cut-off	negative	positive	Total
Cases			
negative	191	10	201
positive	5	32	37
Total	196	42	238
Controls*			
negative	176	2	178
positive	3	1	4
Total	179	3	182

CEA	SNP stratified cut-off		
uniform cut-off	negative	positive	Total
Cases			
negative	202	9	211
positive	4	30	34
Total	206	39	245
Controls*			
negative	254	2	256
positive	0	3	3
Total	254	5	259

* Validation set controls

Table S4. Diagnostic accuracy of SNP-stratified tumor marker combinations

	All PDAC cases n=245 sensitivity^ (95% CI)	PDAC Intact FUT3 n=217 sensitivity^ (95% CI)	Stage1 PDAC n=49 sensitivity^ (95% CI)	Stage1 PDAC (Intact FUT3) n=45 sensitivity^ (95% CI)	Validation set controls n=259 specificity^ (95% CI)	Validation set controls (Intact FUT3) n=237 specificity^ (95% CI)
CA19-9	60.8% (54.4%, 67%)	66.4% (59.6%, 72.6%)	36.7% (23.4%, 51.7%)	40% (25.7%, 55.7%)	98.8% (96.7%, 99.8%)	99.2% (97%, 99.9%)
CA19-9+CEA	64.5% (58.1%, 70.5%)	69.1% (62.5%, 75.2%)	40.8% (27%, 55.8%)	44.4% (29.6%, 60%)	96.9% (94%, 98.6%)	97% (94%, 98.8%)
CA19-9+CA125	64.1% (57.7%, 70.1%)	68.2% (61.6%, 74.3%)	40.8% (27%, 55.8%)	42.2% (27.7%, 57.8%)	97.3% (94.5%, 98.9%)	97.5% (94.6%, 99.1%)
CA19-9+Thrombospondin-2*	63.7% (57.3%, 69.7%)	69.1% (62.5%, 75.2%)	36.7% (23.4%, 51.7%)	40% (25.7%, 55.7%)	96.1% (93%, 98.1%)	96.2% (92.9%, 98.2%)
CA19-9+CEA+CA125	66.1% (59.8%, 72.0%)	70.5% (64%, 76.5%)	42.9% (28.8%, 57.8%)	42.2% (27.7%, 57.8%)	95.4% (92.1%, 97.6%)	95.4% (91.9%, 97.7%)
CA19-9+CEA+CA125+Thrombospondin-2*	68.2% (61.9%, 73.9%)	72.4% (65.9%, 78.2%)	42.9% (28.8%, 57.8%)	42.2% (27.7%, 57.8%)	93.1% (89.2%, 95.8%)	92.8% (88.8%, 95.8%)

* THBS2 (thrombospondin-2) was not measured in most Stage I cases

Stage I by AJCC 8th edition; CI, confidence intervals

^Genotype defined cut-offs