

Group	Patients	Gender	Median age (yrs)	Age range (yrs)	Median bilirubin (mg/mL)	Median CA 19.9 (IU/mL)	CA 19.9 >37 IU/mL	BTC stage $\geq$ T3
BTC	14	7F:7M	73	44-90	19	404	11/14	8/14
H	16	8F:8M	34	23-80	ND	ND	N/A	N/A

**Table S1** Clinical characteristics of external validation set



Mass (m/z)	Average peak area BTC +/- SD	Average peak area benign +/- SD	Fold-change (BTC vs benign)	Wilcoxon test P value	AUROC <sup>†</sup>
887.2	5.23 +/- 7.12	2.93 +/- 4.69	1.78	0.0243	0.74
2082.1	8.97 +/- 3.27	17.12 +/- 15.53	0.52	0.0243	0.74
2210.3	11.21 +/- 3.9	22.92 +/- 19.34	0.49	0.0243	0.75
2554.5	24.45 +/- 17.66	43.06 +/- 27.41	0.57	0.0243	0.74
2903.3	15.44 +/- 8.17	10.06 +/- 5.17	1.53	0.011	0.78
2932.8	79.15 +/- 59.89	138.59 +/- 95.83	0.57	0.0243	0.73
5805.0	39.57 +/- 35.44	21.27 +/- 20.37	1.86	0.0357	0.73

**Table S3** Significant peptide peaks differentiating BTC from benign cases.

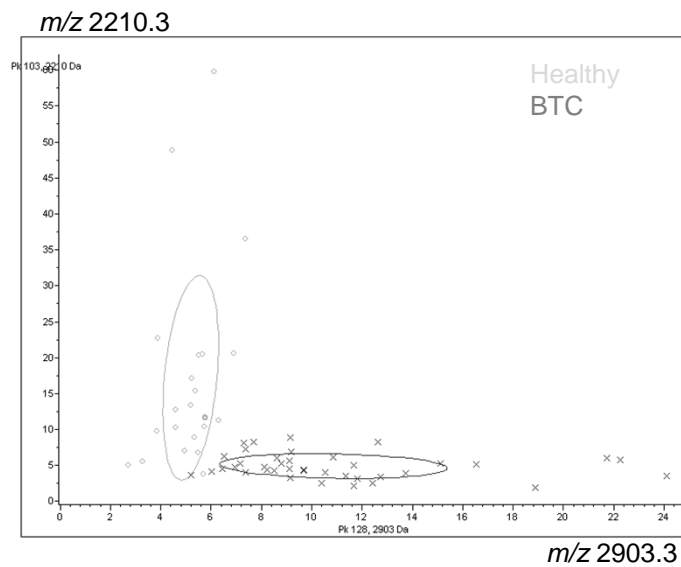
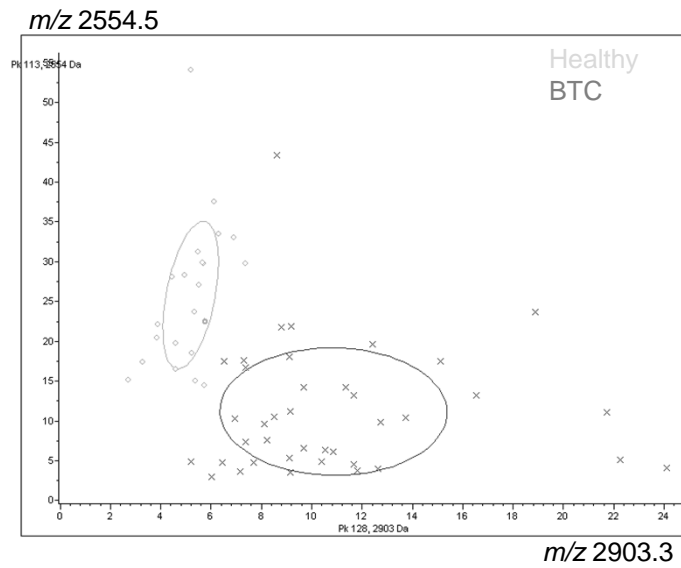
<sup>†</sup>Area under the Receiver Operating Characteristic curve.

Model	Training/test set permutation	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)	Cross-validation (20% leave out) (%)
SVM_9pk_3knn	5	100.00	100.00	100.00	100.00	100.00	94.40%
SVM_5pk_1knn	5	100.00	100.00	100.00	100.00	100.00	93.50%
SVM_7pk_1knn	5	100.00	100.00	100.00	100.00	100.00	92.90%
SVM_9pk_1knn	5	100.00	100.00	100.00	100.00	100.00	92.90%
GA_7pk_1knn	1	100.00	100.00	100.00	100.00	100.00	92.60%
SVM_8pk_1knn	5	100.00	100.00	100.00	100.00	100.00	91.40%
SVM_6pk_1knn	5	100.00	100.00	100.00	100.00	100.00	91.10%
SVM_8pk_1knn	2	100.00	100.00	100.00	100.00	100.00	90.80%
SVM_8pk_3knn	2	100.00	100.00	100.00	100.00	100.00	90.50%
SVM_9pk_3knn	2	100.00	100.00	100.00	100.00	100.00	90.50%
SVM_auto_3knn	2	100.00	100.00	100.00	100.00	100.00	90.50%
SVM_auto_5knn	3	100.00	100.00	100.00	100.00	100.00	90.50%
SVM_3pk_1knn	5	100.00	100.00	100.00	100.00	100.00	89.90%
SVM_7pk_3knn	5	100.00	100.00	100.00	100.00	100.00	89.10%
SVM_auto_3knn	5	100.00	100.00	100.00	100.00	100.00	89.10%
GA_5pk_1knn	5	100.00	100.00	100.00	100.00	100.00	88.40%
GA_7pk_7knn	3	100.00	100.00	100.00	100.00	100.00	88.20%
GA_6pk_1knn	1	100.00	100.00	100.00	100.00	100.00	87.90%
GA_4pk_1knn	5	100.00	100.00	100.00	100.00	100.00	87.90%
GA_9pk_1knn	1	100.00	100.00	100.00	100.00	100.00	87.60%
SVM_9pk_1knn	2	100.00	100.00	100.00	100.00	100.00	87.60%
GA_4pk_1knn	2	100.00	100.00	100.00	100.00	100.00	85.80%
SVM_2pk_1knn	5	100.00	100.00	100.00	100.00	100.00	83.50%
GA_8pk_3knn	2	100.00	100.00	100.00	100.00	100.00	79.40%
SNN	5	100.00	100.00	100.00	100.00	100.00	73.80%

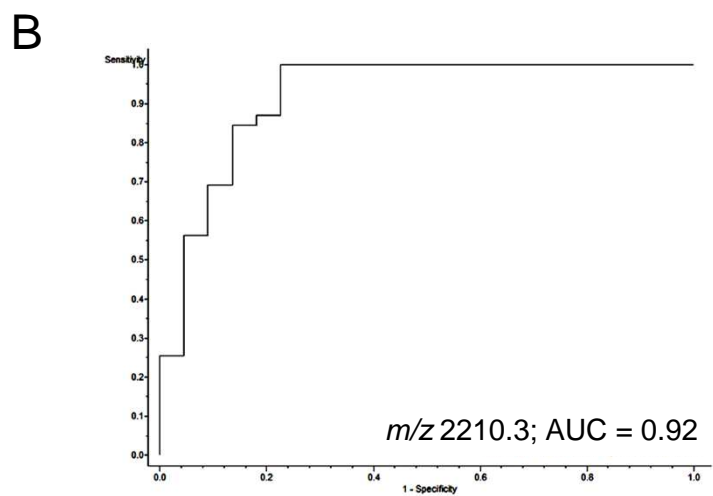
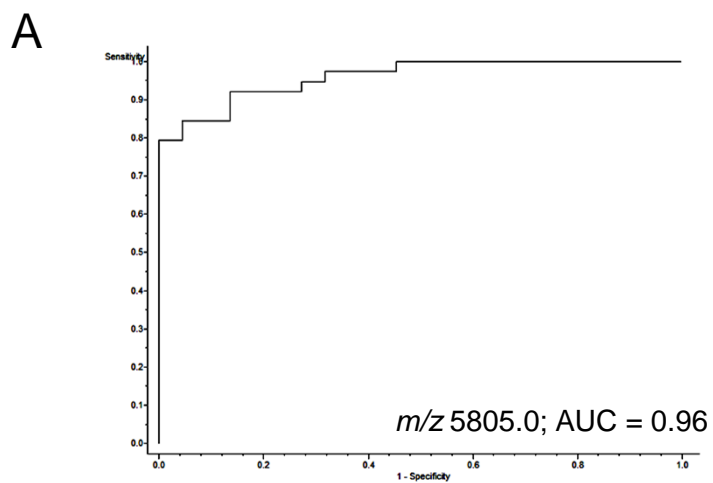
**Table S4** Best-performing predictive models for determining BTC versus healthy. A SVM 9 peak model was the best performing on cross-validation and for test set classification (shaded grey).

Ave Mass by MALDI-TOF (m/z)	Calculated average mass (Da)	Calculated average mass [M+H] <sup>+</sup>	Calculated monoisotopic mass (Da)	Name	Sequence	Identification
887.2						no
1020.2	1020.11	1021.12	1019.50	Fibrinopeptide A	DFLAEGGGVR	yes; Villanueva et al; Tiss et al
1263.7	1263.33	1264.34	1262.59	Fibrinopeptide A	GEGDFLAEGGGVR	yes; Villanueva et al; Tiss et al
1350.8	1350.41	1351.42	1349.62	Fibrinopeptide A	SGEGDFLAEGGGVR	yes; Villanueva et al; Tiss et al
1363.4						no
2082.1	2086.07	2087.08	2084.89	Fibrinogen alpha	GGSTSYGTGSETESPRNPSSAG	Koomen et al
2210.3	2209.33	2210.34	2208.05	HMW kininogen	KHNLGHGHKHERDQGHGHQ	Villanueva et al; Tiss et al
2554.5	2553.59	2554.60	2552.09	Fibrinogen alpha	SSSYSKQFTSSTSYNRGDSTFES	Villanueva et al; Tiss et al
2624.6	2624.93	2625.94	2623.40	Fibrinogen alpha	LVTSKGDKELRTGKEKVTSGSTTTT	Koomen et al
2903.3						no
2932.8	2932.02	2933.03	2930.28	Fibrinogen alpha	SSSYSKQFTSSTSYNRGDSTFESKSY	Villanueva et al; Tiss et al
5064.8						no
5805.0	5805.09	5806.10	5801.63	Fibrinogen alpha	SSSYSKQFTSSTSYNRGDSTFESKSYK MADEAGSEADHEGTHSTKRGHAKSRP	yes

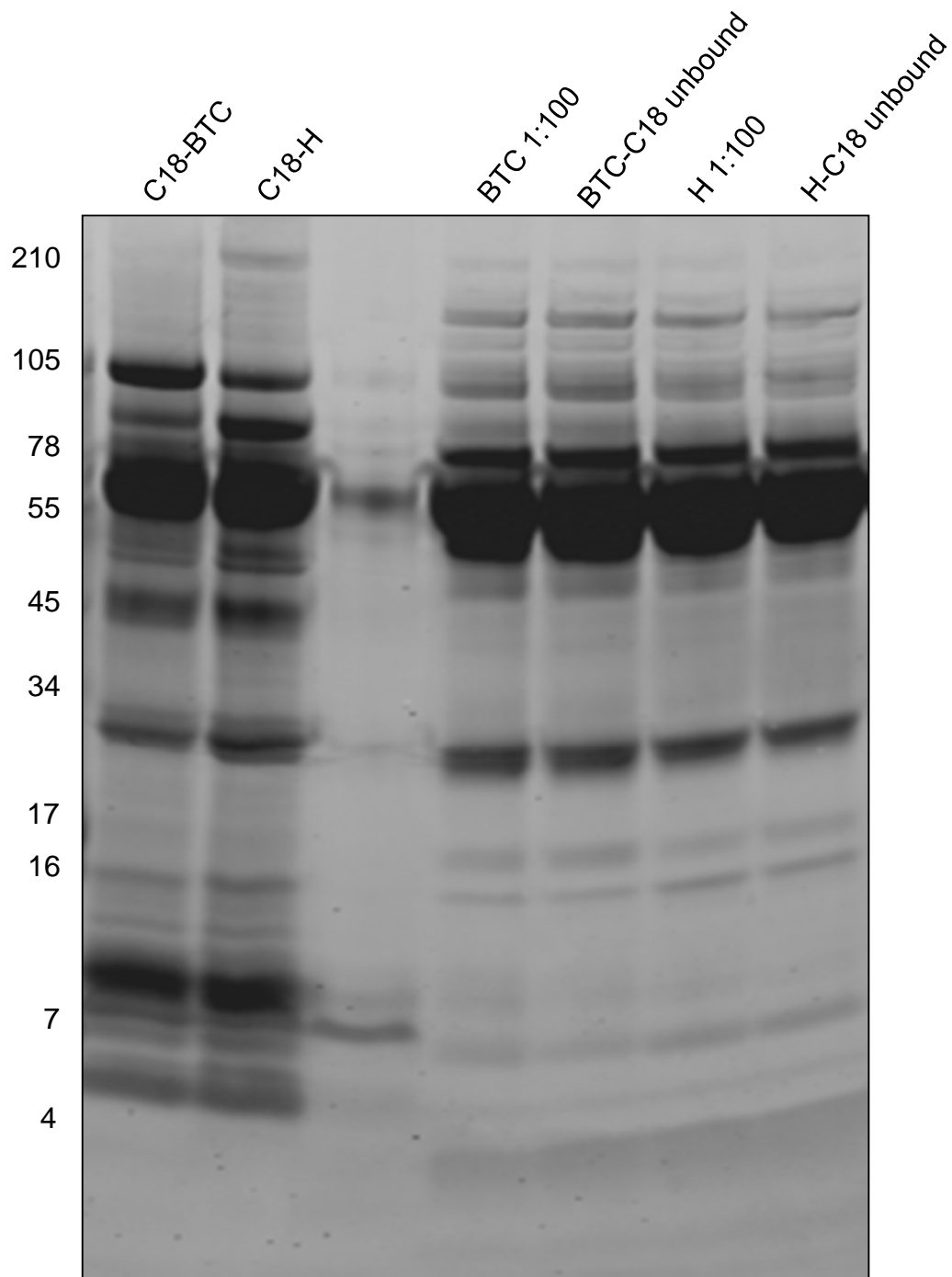
**Table S5** List of identified and predicted peak identifications. C18-extracted and gel separated serum samples were subjected to high-mass accuracy LC-MS/MS as described in Materials and Methods. Predicted identifications were made from published lists of MALDI-TOF and LC-MS/MS peak identifications, including our own previous work (Ref 17; Tiss et al.).



**Figure S1** 2D scatter plots of peak intensities for **A.** *m/z* 2554.5 vs *m/z* 2903.3 and **B.** *m/z* 2210.3 vs *m/z* 2903.3 showing separation of BTC and healthy samples.



**Figure S2** Receiver Operating Characteristics (ROC) curve analysis. **A.** The peak at  $m/z$  5805.0 gave an area under the curve (AUC) of 0.96 for BTC versus healthy samples. **B.** The peak at  $m/z$  2210.3 gave an AUC of 0.92 for BTC versus healthy samples.



**Figure S3** 1D-Tris-tricine (10-20%) gel of C18 bead-extracted sera, unbound material and crude sera. Protein loads (~15 µg) were C18-extracted BTC sample pool (C18-BTC), C18-extracted healthy sample pool (C18-H), crude BTC serum (BTC 1:100), crude healthy serum (H 1:100;), unbound BTC material (BTC-C18 unbound) and unbound healthy material (H-C18 unbound). Bands below 20 kDa were cut from the gel, proteins/peptides passively eluted and analysed by LC-MS/MS.