On-line supplement to: "N. Fananapazir, M. Li, D. Spentzos, C.F. Aliferis; Formative Evaluation of a Prototype System for Automated Analysis of Mass Spectrometry Data"

Papers that explore the role of MS data in clinical applications

Domains involve a variety of tissue types – blood serum, tissue biopsy, nipple aspirate fluid, pancreatic juice – in the analysis of a variety of cancers – ovarian[1,2,3,4], prostate[5,6,7,8], renal[9,10], breast[11,12], head and neck[13], lung[14,15,16], laryngeal[17], hepatic[18], cervical[19], pancreatic[20,21], colorectal[22], bladder[23] – and non-cancers – hepatitis[18], and cerebrovascular accidents[24].

Table 1: Current methods used in MS data analysis

Reference	Pub. Date	Domain	Number of Samples				Study Design			Metric
			Healthy	*Diseased or Benign (non-Cancer)*	*Cancer*	*Total*	*Overalll Study design*[+]	*Reported Pre-processing Steps*[++]	*Classifier*[+++]	
Petricoin[1]	02/2002	ovarian cancer	100	16	100	216	1-fold		GA	sensitivity/specificity
Li[12]	05/2002	breast cancer	41	25	103	169	100-fold	NR P	multivariate	sensitivity/specificity.
Adam[8]	07/2002	prostate cancer	82	77	167	326	1-fold	NRBPA	DT	sensitivity/specificity
Qu[6]	07/2002	prostate cancer	96	92	197	386	10-fold	NR P	ROC-analysis DT (boosted)	sensitivity/specificity.
Vlahou[4]	02/2003	ovarian cancer	95	0	44	139	10-fold	NRPA	DT	accuracy
Yanagisawa[15]	08/2003	lung cancer	14	0	79	93	LOOCV	RBPA	WFCCM	accuracy
Hilario[16]	09/2003	lung cancer	17	0	24	41	10-fold	N PBA	DT KNN MLP Naïve Bayes	accuracy
Kozak[3]	10/2003	ovarian cancer	56	19	109	184	1-fold	NR P	multivariate	sensitivity/specificity ROC accuracy
Won[10]	12/2003	renal cancer	6	15	15	36	0-fold	RBPA	DT	sensitivity/specificity accuracy
Koopmann[20]	02/2004	pancreatic cancer	60	60	60	180	30-fold	NRBP	multivariate	sensitivity/specificity ROC
Wadsworth[13]	03/2004	head and neck cancer	102	0	99	201	1-fold	NR P	DT	sensitivity/specificity
Zhu[18]	08/2004	liver cancer	25	25	20	70	1-fold	PA	DT	sensitivity/specificity PPV
Wong[19]	08/2004	cervical cancer	27	0	35	62	1-fold	NRBPA	ROC-analysis	sensitivity/specificity positive/negative PV
Prados[24]	08/2004	cerebrovascular accidents	0	21	21	42	10-fold	NRBPA	SVM KNN MLP	sensitivity/specificity
Vlahou[23]	08/2004	bladder cancer	33	92	105	230	1-fold	N PA	DT	sensitivity/specificity
Yu[22]	11/2004	colorectal cancer	92	35	55	182	10-fold	NR P	SVM NN	sensitivity/specificity

[+] **Overall study design key:** n-fold: n-fold cross-validation, LOOCV: leave-one-out cross validation
[++] **Pre-processing key:** N: normalization, R: range restriction, B: baseline subtraction, P: Peak detection and/or binning, A: Peak alignment
[+++] **Classifier key:** SVM: support vector machine, NN: neural network, GA: genetic algorithm, DT: decision tree, MLP: multi-layer perception, KNN: K-nearest neighbour, WFCCM: weighted flexible compound covariate method

Table 2: Report of Performance of Generated Models

		Prior familiarity with FAST-AIMS and/or MS	Computer time to generate model[++]	User time	Strategies Employed[+++]	Estimated performance (ROC)	Actual Performance (ROC)
FAST-AIMS users	User 1	Y	8 hours	< 30 minutes	LOOCV BC, PD, PA AF, RFE, HITON SVM-gauss	0.810	0.802
	User 2	Y	9 hours		10-fold BC,PD, PA AF, HITON SVM-poly	0.773	0.779
	User 3	Y	19 hours		10-fold BC,PD, PA AF, HITON SVM-poly	0.760	0.773
	User 4	Y	3 hours		10-fold HITON SVM-poly	0.717	0.773
	User 5	N	55 hours		10-fold BC,PD, PA AF, RFE, HITON SVM-gauss	0.786	0.777
	User 6	N	22 hours		LOOCV BC,PD, PA AF, RFE, HITON SVM-poly	0.789	0.773
Expert Biostatistician[+]				7 hours	UDWT, BC, WFCCM	0.808	0.811

[+] *Model developed independently of FAST-AIMS*
[++] *For FAST-AIMS users, time to generate model is computation time (not user time). User time was < 30 minutes in all cases.*
[+++] ***Strategies employed key:*** *n-fold: n-fold cross validation, LOOCV: leave-one-out cross-validation, BC: baseline correction (Coombes), PD: peak detection (Coombes), peak alignment (Coombes), AF: all features, RFE: recursive feature elimination, SVM-poly: support vector machine (polynomial kernel), SVM-gauss: support vector machine (gaussian kernel), UDWT: undecimated discrete wavelet transformation, WFCCM: weighted flexible compound covariate method*

Additional References

[1] Petricoin EF, Ardekani AM, Hitt BA, et al. *Use of proteomic patterns in serum to identify ovarian cancer*. Lancet 2002; 359: 572-577
[2] Conrads TP, Fusaro VA, Ross S, et al. *High-resolution serum proteomic features for ovarian cancer detection*. Endocr Relat Cancer. 2004 Jun;11(2):163-78
[3] Kozak KR, Amneus MW, Pusey SM, et al. *Identification of biomarkers for ovarian cancer using strong anion-exchange Proteinchips: Potential use in diagnosis and prognosis*. Proc. Natl. Acad. Sci 2003 100, 12343 –12348
[4] Vlahou A, Schorge JO, Gregory BW, Coleman RL. *Diagnosis of Ovarian Cancer Using Decision Tree Classification of Mass Spectral Data*. J Biomed Biotechnol. 2003;2003(5):308-314.
[5] Petricoin EF, et al. *Serum proteomic patterns for detection of prostate cancer*. J Natl Cancer Inst. 2002 Oct 16;94(20):1576-8
[6] Qu Y, Adam BL, Yasui Y, et al. *Boosted Decision Tree Analysis of Surface-enhanced Laser Desorption/Ionization Mass Spectral Serum Profiles Discriminates Prostate Cancer from Noncancer Patients.* Clin Chem. 2002 Oct;48(10):1835-43.
[7] Banez LL, et al. *Diagnostic potential of serum proteomic patterns in prostate cancer*. J Urol. 2003 Aug;170(2 Pt 1):442-6.
[8] Adam BL, Qu Y, Davis JW, et al. *Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men*. Cancer Res. 2002 Jul 1;62(13):3609-14.

[9] Junker K, Gneist J, Melle C, et al. *Identification of protein pattern in kidney cancer using ProteinChip(R) arrays and bioinformatics*. Int J Mol Med. 2005 Feb;15(2):285-90.
[10] Won Y, Song HJ, Kang TW, et al. *Pattern analysis of serum proteome distinguishes renal cell carcinoma from other urologic diseases and healthy persons.* Proteomics 2003, 2:2310-2316.
[11] Coombes, KR, Fritsche Jr. HA, Clarke C, et al. *Quality Control and Peak Finding for Proteomics Data Collected from Nipple Aspirate Fluid by Surface-Enhanced Laser Desorption and Ionization.* Clinical Chemistry 2003, 49:10, 1615-1623.
[12] Li J, Zhang Z, Rosenzweig J, Wang YY, Chana DW; *Proteomics and Bioinformatics Approaches for Identification of Serum Biomarkers to Detect Breast Cancer.* Clinical Chemistry. 2002;48:1296-1304.
[13] Wadsworth JT, Somers KD, Cazares LH, et al. *Serum protein profiles to identify head and neck cancer.* Clin Cancer Res. 2004 Mar 1;10(5):1625-32.
[14] Xiao X, Liu D, Tang Y, et al. *Development of proteomic patterns for detecting lung cancer*. Dis Markers. 2003-2004;19(1):33-9.
[15] Yanagisawa K, Shyr Y, Xu BJ, et al. Nadaf S, Moore JH, Caprioli RM, Carbone DP. *Proteomic patterns of tumour subsets in non-small-cell lung cancer.* Lancet. 2003 Aug 9;362(9382):433-9.
[16] Hilario M, Kalousis A, Muller M, Pellegrini C. *Machine learning approaches to lung cancer prediction from mass spectra.* Proteomics 2003, 3, 1716-1719.
[17] Xiao X, Zhao X, Liu J, Guo F, Liu D, He D. *Discovery of laryngeal carcinoma by serum proteomic pattern analysis*. Sci China C Life Sci. 2004 Jun;47(3):219-23.
[18] Zhu XD, Zhang WH, Li CL, Xu Y, Liang WJ, Tien P. *New serum biomarkers for detection of HBV-induced liver cirrhosis using SELDI protein chip technology*. World J Gastroenterol. 2004 Aug 15;10(16):2327-9.
[19] Wong YF, Cheung TH, Lo KW, et al. *Protein profiling of cervical cancer by protein-biochips: proteomic scoring to discriminate cervical cancer from normal cervix*. Cancer Lett. 2004 Aug 10;211(2):227-34.
[20] Koopmann J, Zhang Z, White N, et al. *Serum diagnosis of pancreatic adenocarcinoma using surface-enhanced laser desorption and ionization mass spectrometry*. Clin Cancer Res. 2004 Feb 1;10(3):860-8.
[21] Rosty C, Christa L, Kuzdal S, et al. *Identification of Hepatocarcinoma-Intestine-Pancreas /Pancreatitis -associated Protein I as a Biomarker for Pancreatic Ductal Adenocarcinoma by Protein Biochip Technology.* Cancer Research 2002, 62, 1868-1875.
[22] Yu JK, Chen YD, Zheng S. *An integrated approach to the detection of colorectal cancer utilizing proteomics and bioinformatics*. World J Gastroenterol. 2004 Nov 1;10(21):3127-31.
[23] Vlahou A, et al. *Protein profiling in urine for the diagnosis of bladder cancer*. Clin Chem. 2004 Aug;50(8):1438-41.
[24] Prados J, Kalousis A, Sanchez JC, Allard L, Carrette O, Hilario M *Mining mass spectra for diagnosis and biomarker discovery of cerebral accidents*. Proteomics. 2004 Aug;4(8):2320-32.