Supporting Information for

Diagnosis of early stage ovarian cancer by ^{1}H NMR metabonomics of serum explored by use of a micro-flow NMR probe

Erwin Garcia,[†] Chris Andrews,[‡] Jia Hua,[†] Hyung L. Kim,[‖] Dinesh K. Sukumaran,[†] Thomas Szyperski,[†,*] and Kunle Odunsi[§,*]

Departments of Chemistry and Biostatistics, The State University of New York at Buffalo, and the Departments of Urologic Oncology and Gynecological Oncology, Roswell Park Cancer Institute, Buffalo, NY, USA

Clinical Characteristics of EOC / RCC patients

EOC Training / Test Set: (n = 252)

Median age: 49 years (range: 25, 67). The histological characteristics of the 120 cancer cases were endometrioid (n = 45), serous (n = 21), clear cell (n = 17), mucinous (n = 26), and mixed (n = 11). Patient stages are I (n=86) and II (n=33). Patient grades are 1 (n = 43), 2 (n = 38), 3 (n = 26), and 4 (n = 4).

EOC Validation Set: (n = 100)

Median age: 49 years (range: 27 to 63). Patient stages are I (n = 35) and II (n = 15).

RCC Set: (n = 30)

Median age: 66 years (range 34 to 92). Patients had clear cell RCC (n = 26; 87%) and papillary RCC (n = 4; 13%).

Additional Plots of NMR spectra

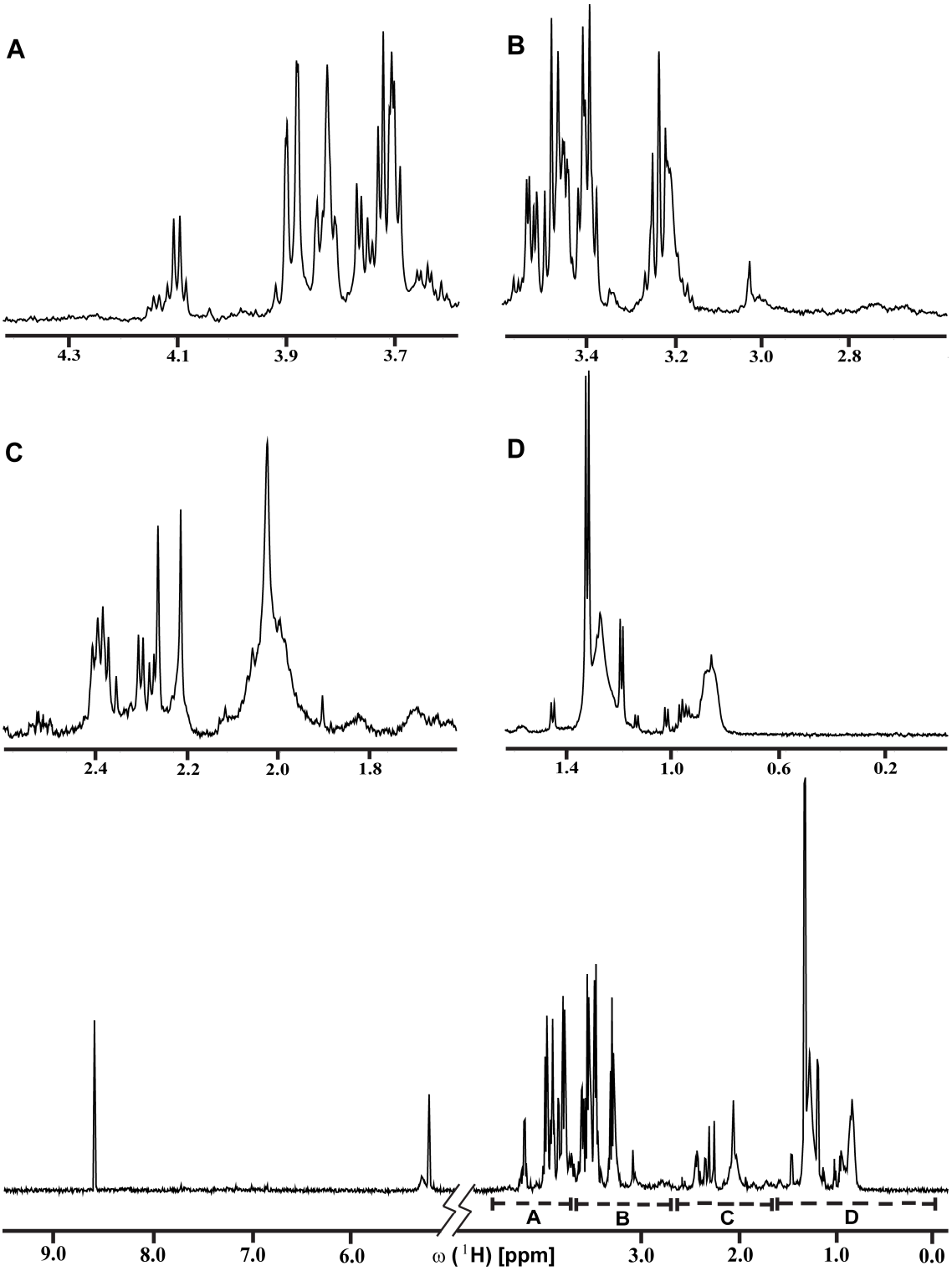

Figure S1. Detailed expansions of different spectral regions of the representative 1D CPMG spectrum of Fig. 1.

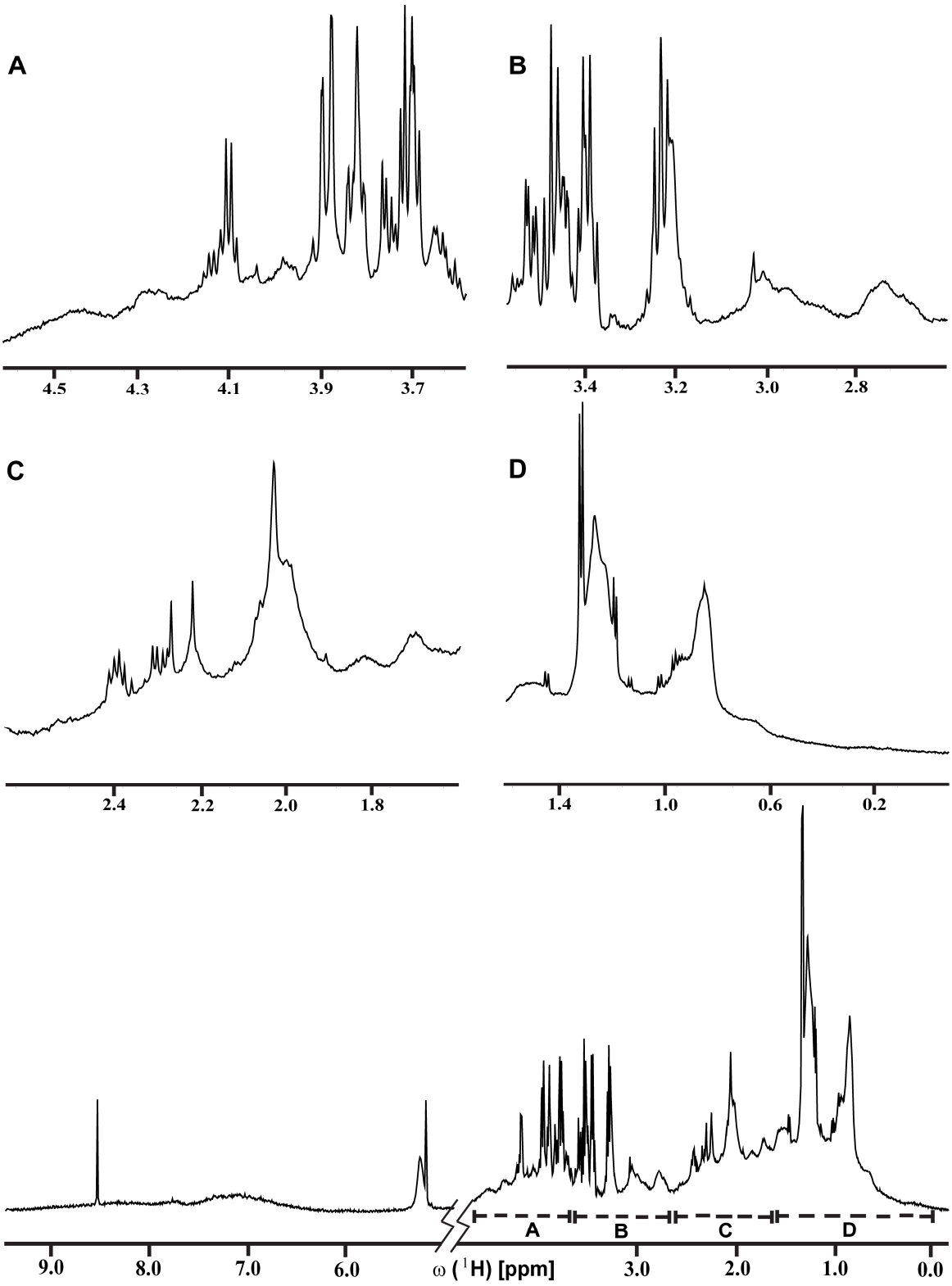

Figure S2. Detailed expansions of different spectral regions of the representative 1D NOESY spectrum of Fig. 1.

Multi-variate data analysis

Analysis of spectra recorded for Renal Cell Cancer (RCC) samples

NMR spectra were acquired for 30 specimens obtained from newly diagnosed female RCC patients and processed as described above for the EOC study. The predictive EOC model was applied. Seven specimens of 30 (23%) resulted in positive tests. This rate from the RCC study is somewhat higher than but not statistically significantly different (Fisher p = 0.12) from the false positive rate in the EOC study (10 of 94, 11%, Combined Test and Validation Sets). Specifically, the RCC positive rate (23%) is very similar to the false positive rate in the Test Set (9 of 44, 20%, p = 0.78) but (like the false positive rate for the Test Set) higher than for the Validation Set alone (1 of 50, 2%, p = 0.004).

Relationship between Sensitivity (Sns), Specificity (Spc), Prevalence (Prv), and Positive Predictive Value (PPV)

Bayes Rule, a simple equation regarding conditional probabilities, relates these four quantities so that one can be determined from the other three: $PPV = Spc * Prv / (Spc * Prv + (1-Sns) * (1 - Prv))$. The sensitivity (i.e., the probability of a positive test result given a sample from an early stage EOC patient) and the specificity (i.e., the probability of a negative test result given a sample from a healthy control) can be directly estimated from a case-control study [Pepe, M, The Statistical Evaluation of Medical Tests for Classification and Prediction, Oxford University Press, 2003]. To compute the PPV it is necessary to know also the prevalence of the disease. Supplementary Table S1 displays the PPV for a variety of combinations of sensitivity and specificity and three different risk populations. Standard confidence intervals for the sensitivity and specificity can be transformed to a confidence interval for PPV via the multivariate delta method [Pepe, M, The Statistical Evaluation of Medical Tests for Classification and Prediction, Oxford University Press, 2003]. In a population at 20-fold risk of EOC (*i.e.* slightly less than the risk of BRCA2 carriers) over the general population (1/100) a test with 80% sensitivity and 90% specificity yields a PPV of 7.5% i.e. 13 positive screens per EOC. At even higher risks e.g. 3/100 (i.e. 67-fold over the general population, slightly less than BRCA1 carriers), even a test with 50% sensitivity and 86% specificity has a 10% PPV.

Tables

Table S1. Operating Characteristics of predictive models built with (a) CPMG bin arrays ('CPMG), (b) NOESY bin arrays ('NOESY') alone, and (c) concatenated CPMG and NOESY bin arrays ('joint'). The area under the ROC Curve (AUC) measures the quality of predictive model based on the p-EOC computed for each spectrum. AUC values are similar for the three predictive models with the joint model being slightly superior when compared with the separate models for the Test Set, the Validation Set, and the Test and Validation Sets combined. Alternatively we can dichotomize p-EOC at an arbitrary 'cut-point' to provide a binary ('+'/'-') decision rule and compute the specificity (probability of correctly identifying a healthy control) and sensitivity (probability of correctly identifying an early stage EOC). For this table the prevalence of disease in each set was used as the cut-point (40/88 in the Test Set; 50/100 in the Validation Set).

TEST SET

	(a) CPMG		(b) NOESY		(c) Joint	
	Healthy Control	Early Stage EOC	Healthy Control	Early Stage EOC	Healthy Control	Early Stage EOC
AUC (95% CI)	.715 (.600,.831)		.763 (.658,.867)		.796 (.696,.897)	
Healthy Control	36	19	33	13	35	15
Early Stage EOC	8	21	11	27	9	25
Specificity (95% CI)	82% (67%,92%)		75% (60%,87%)		80% (65%,90%)	
Sensitivity (95% CI)		53% (36%,68%)		68% (51%,81%)		63% (46%,77%)

VALIDATION SET

	(a) CPMG		(b) NOESY		(c) Joint	
	Healthy Control	Early Stage EOC	Healthy Control	Early Stage EOC	Healthy Control	Early Stage EOC
AUC (95% CI)	.905 (.831,.979)		.934 (.885,.983)		.949 (.896,1.000)	
Healthy Control	48	16	50	17	49	13
Early Stage EOC	2	34	0	33	1	37
Specificity (95% CI)	96% (86%,100%)		100% (93%,100%)		98% (89%,100%)	
Sensitivity (95% CI)		68% (53%,80%)		66% (51%,79%)		74% (60%,85%)

COMBINED SETS

	(a) CPMG		(b) NOESY		(c) Joint	
	Healthy Control	Early Stage EOC	Healthy Control	Early Stage EOC	Healthy Control	Early Stage EOC
AUC (95% CI)	.820 (.753,.888)		.852 (.795,.909)		.880 (.825,.935)	
Healthy Control	84	35	83	30	84	28
Early Stage EOC	10	55	11	60	10	62
Specificity (95% CI)	89% (81%,95%)		88% (80%,96%)		89% (81%,95%)	
Sensitivity (95% CI)		61% (50%,71%)		67% (56%,76%)		69% (58%,78%)

Table S2. Positive predictive value (PPV) as a function of incidence, specificity and sensitivity. PPVs below the solid line in the table are above the threshold of 10%, which is considered a lower bound for clinical applications.[5]

Positive Predictive Value

Incidence Rate (per 100,000)		45 General Population			100 High Risk			3000 Higher Risk		
	Sensitivity	**50%**	**80%**	**100%**	**50%**	**80%**	**100%**	**50%**	**80%**	**100%**
Specificity	**80%**	0.1%	0.2%	0.2%	0.2%	0.4%	0.5%	7.2%	11.0%	13.4%
	90%	0.2%	0.4%	0.4%	0.5%	0.8%	1.0%	13.4%	19.8%	23.6%
	95%	0.4%	0.7%	0.9%	1.0%	1.6%	2.0%	23.6%	33.1%	38.2%
	97%	0.7%	1.2%	1.5%	1.6%	2.6%	3.2%	34.0%	45.2%	50.8%
	99%	2.2%	3.5%	4.3%	4.8%	7.4%	9.1%	60.7%	71.2%	75.6%
	99.6%	5.3%	8.3%	10.1%	11.1%	16.7%	20.0%	79.4%	86.1%	88.5%
	99.8%	10.1%	15.3%	18.4%	20.0%	28.6%	33.4%	88.5%	92.5%	93.9%