Supplementary Text and Figures

Table S1. General information on patients and their cancer types whose plasma was used in this study.	p. S2
Table S2. Classification results obtained from two developed linear and nonlinear models for six groups of samples.	p. S3
Table S3. Correlation coefficient of CPANN weight map for each variable in the six classes.	p. S4
Fig. S1. Classification of identified coronas by sensor array elements according to their physiological functions, including (A) acute phase, (B) coagulation, (C) immunoglobulins, (D) lipoproteins, (E) tissue leakage, and (F) other plasma proteins, in human plasma of healthy subjects and patients having different types of cancers.	
Fig. S1A. Acute phase.	p. S6
Fig. S1B. Coagulation.	p. S7
Fig. S1C. Immunoglobulins.	p. S8
Fig. S1D. Lipoproteins.	p. S9
Fig. S1E. Tissue Leakage.	p. S10
Fig. S1F. Complement.	p. S11
Fig. S1G. Other Plasma Proteins.	p. S12
Fig. S2. Schematic representation of study outline. Informative variable selection and classification model building.	p. S13
Fig. S3. ROC (receiver operating characteristic) plots derived from PLS-DA and confusion matrix drived from CPANN (counter propagation artificial neural network).	p. S14
Fig. S4. Data analysis using CPANN (counter propagation artificial neural network).	p. S15
Excel Data Files, Six	

PATIENT LABEL	AGE	GENDER	CANCER STAGE	CANCER STAGE CATEGORY
PANCREAS 1	81	М	TNM: cT4 N+M+	Advanced
PANCREAS 2	76	М	TNM: cT2 N+	Moderate
PANCREAS 3	60	F	TNM: cT4 N+	Advanced
PANCREAS 4	75	М	TNM: cT3 N+M+	Advanced
PANCREAS 5	61	F	TNM: cT3 N+M+	Advanced
LUNG 1	45	F	TNM: TIA N0 M0	Early
LUNG 2	24	М	TNM: TIA N0 M0	Early
LUNG 3	44	F	TNM: TIA N0 M0	Early
LUNG 4	48	F	TNM: TIA N0 M0	Early
LUNG 5	40	М	TNM: TIA N0 M0	Early
MYELOMA 1	41	F	Onset II (52% Plasmacells in Bone Marrow; Monoclonal IgG-k)	Moderate
MYELOMA 2	60	F	Onset I (28% Plasmacells in Bone Marrow; Monoclonal IgG-k)	Early
MYELOMA 3	57	F	Onset I (8% Plasmacells in Bone Marrow; Monoclonal IgG-k)	Early
MYELOMA 4	63	М	Onset I (9% Plasmacells in Bone Marrow; Monoclonal IgG-k)	Early
MYELOMA 5	75	М	Onset II (42% Plasmacells in Bone Marrow; Monoclonal IgG-k)	Moderate
GLIOBLASTOMA 1	58	М	WHO 4	Advanced
GLIOBLASTOMA 2	75	М	WHO 4	Advanced
GLIOBLASTOMA 3	76	М	WHO 4	Advanced
GLIOBLASTOMA 4	73	М	WHO 4	Advanced
GLIOBLASTOMA 5	76	F	WHO 4	Advanced
MENINGIOMA 1	82	F	WHO 2	Moderate
MENINGIOMA 2	50	F	WHO 2	Moderate
MENINGIOMA 3	80	М	WHO 2	Moderate
MENINGIOMA 4	67	М	WHO 2	Moderate
MENINGIOMA 5	64	М	WHO 2	Moderate

Table S1. General information on patients and their cancer types whose plasma was used in this study.

Table S2. Classification results obtained from two developed linear and nonlinear models for six groups of samples and also three groups of cohort samples (10-fold cross validation is repeated 10 times and the reported model performance statistics are averaged).

Models/ classification parameters		Fresh samples		Cohort samples					
		PLS-DA, LV=5	CPANN (8*8)	PLS-DA, 69 variables LV=3		CPANN 69 variables (8*8)	PLS-DA 8 variables LV=2	CPANN 8 variables (8*8)	
	1	1.00	1.00	1					
	2	0.96	0.984		0.591	0.57	1.00	1.00	
Specificity	3	1.00	0.996	2					
specificity	4	1.00	1.00		0.760	0.86	1.00	1.00	
	5	1.00	1.00	3					
	6	1.00	1.00		1.00	0.96	1.00	1.00	
	1	1.00	0.98	1					
	2	1.00	0.98		0.751	0.72	1.00	1.00	
Sensitivity	3	1.00	0.94	2					
	4	1.00	1.00		0.850	0.32	1.00	1.00	
	5	1.00	1.00	. 3					
	6	1.00	1.00		0.540	0.74	1.00	1.00	
	1	0.00	0.033	2					
	2	0.03	0.016		0.329	0.38	0.00	0.00	
Class error	3	0.00	0.013						
	4	0.00	0.00		0.201	0.32	0.00	0.00	
	5	0.00	0.00	3					
	6	0.00	0.00		0.131	0.11	0.00	0.00	

1) Control 2) Glioblastoma 3) Meningioma 4) Myeloma 5) Pancreas 6) Lung cancers

Cohort samples: 1) Brain 2) Lung 3) Pancreas

LV (Latent variables): Partial Least Square defines a low dimensional space based on small number of orthogonal factor which called latent factors or variables. Each factor is a linear combination of the original variables. PLS extracts these latent factors, accounting for as much of the variation in data space as possible while modeling the responses properly.

CV (Cross Validation): Cross-validation is a method of assessing generalization performance of statistical models like classifiers. One of the most popular approaches to do CV is the holdout method. This could be done by splitting the original dataset into complementary subsets and performing the analysis using one subset (training set), and validating the results using the leavedout subset (validation set).

Sensitivity= TP/(TP+FN),

Specificity= TN/(TN+FP)

Class error= 1-accuracy = (FP+FN)/(TP+TN+FP+FN)

TP: true positive; TN: true negative; FP: false positive; FN: false negative

Variable	CC(1),	CC(2)	CC(3)	CC(4)	CC(5)	CC(6)
ID	Control	Glioblastoma	Meningioma	Myeloma	Pancreas	Lung
1	0.60	-0.31	-0.39	-0.37	-0.13	0.46
2	-0.23	-0.13	-0.22	-0.18	0.92	-0.11
3	-0.20	-0.10	-0.27	-0.15	0.93	-0.16
4	0.38	-0.06	-0.21	-0.72	-0.03	0.58
5	-0.38	-0.23	-0.36	0.33	0.05	0.70
6	-0.01	-0.36	-0.58	0.42	0.01	0.56
7	-0.37	-0.32	-0.39	0.66	0.02	0.50
8	-0.15	-0.19	-0.32	0.02	0.94	-0.25
9	-0.45	0.11	-0.13	0.87	-0.22	-0.03
10	-0.51	-0.22	-0.15	0.55	0.57	-0.13
11	-0.33	-0.15	-0.24	0.65	-0.45	0.61
12	-0.34	-0.24	-0.43	0.36	0.02	0.72
13	0.18	0.40	0.38	-0.55	0.02	-0.47
14	-0.47	-0.04	-0.07	0.72	-0.49	0.47
15	-0.33	-0.18	-0.21	0.64	-0.43	0.60
16	0.15	-0.28	-0.44	0.16	-0.32	0.72
17	-0.09	-0.26	-0.32	0.85	-0.06	-0.08
18	-0.20	-0.03	-0.19	-0.21	0.89	-0.20
19	-0.12	-0.09	-0.30	-0.18	0.98	-0.24
20	-0.16	-0.16	-0.25	-0.13	-0.16	0.91
21	-0.37	-0.05	-0.33	0.55	-0.34	0.68
22	-0.03	-0.16	-0.32	-0.20	-0.13	0.87
24	-0.18	-0.13	-0.26	-0.14	0.96	-0.20
25	-0.19	-0.10	-0.25	-0.20	0.99	-0.20
26	-0.12	-0.15	-0.25	-0.09	-0.17	0.82
27	0.90	-0.24	-0.29	-0.21	-0.17	-0.18
28	-0.16	-0.13	-0.22	-0.12	-0.16	0.83
29	-0.29	-0.11	0.86	-0.08	-0.23	-0.23
30	-0.18	0.69	-0.02	-0.26	-0.05	-0.02
31	-0.37	0.33	0.67	-0.17	0.16	-0.59
32	0.56	-0.01	-0.01	-0.61	-0.44	0.39
33	-0.60	0.65	0.57	-0.08	-0.02	-0.37
34	-0.45	0.37	0.74	-0.17	0.07	-0.51
35	-0.45	0.50	-0.08	0.52	0.05	-0.34
36	-0.47	-0.04	0.11	0.90	-0.03	-0.37
37	-0.47	-0.01	0.42	0.74	-0.34	-0.28
38	0.41	-0.35	0.18	-0.31	-0.51	0.40
39	-0.23	-0.08	0.01	0.78	-0.08	-0.37
40	0.29	-0.32	-0.37	0.75	-0.20	-0.19
41	-0.57	0.14	0.78	0.14	-0.03	-0.43
42	0.42	0.26	-0.38	-0.45	-0.31	0.46
43	-0.28	0.03	0.59	-0.05	0.37	-0.68
44	-0.15	0.34	0.66	-0.62	0.08	-0.33
45	0.81	-0.04	-0.43	-0.35	-0.22	0.11
46	-0.16	0.04	0.80	-0.41	0.16	-0.51

Table S3. Correlation coefficient of CPANN weight map for each variable in the six classes.

47	0.68	-0.19	0.04	-0.41	-0.44	0.13
48	-0.40	-0.10	-0.32	0.03	0.69	0.22
49	-0.45	0.82	0.12	-0.07	-0.05	-0.16
50	0.87	-0.35	-0.21	-0.05	-0.36	-0.12
51	0.47	0.01	-0.56	-0.14	-0.30	0.50
52	-0.21	-0.11	-0.05	-0.23	0.87	-0.25
53	-0.63	0.12	0.81	-0.12	0.05	-0.19
54	-0.30	-0.13	-0.10	0.98	-0.21	-0.16
55	0.59	-0.05	-0.10	-0.61	-0.36	0.40
56	-0.19	-0.21	-0.32	-0.09	0.09	0.77
57	0.53	-0.35	0.42	-0.44	-0.24	-0.16
58	-0.14	-0.25	-0.17	-0.20	-0.12	0.89
59	-0.37	0.03	0.84	-0.31	0.14	-0.37
60	0.42	-0.11	-0.39	-0.32	-0.32	0.65
61	-0.33	0.66	0.07	-0.18	-0.26	0.22
62	-0.45	0.10	0.78	-0.20	0.03	-0.26
63	-0.18	-0.08	-0.24	-0.18	0.93	-0.19
64	-0.22	0.76	-0.12	-0.12	-0.15	0.05
65	-0.12	-0.05	-0.26	-0.15	-0.21	0.85
66	-0.18	-0.09	-0.25	-0.19	0.98	-0.20
67	0.95	-0.24	-0.35	-0.28	-0.10	-0.17
68	-0.16	0.58	-0.15	-0.16	0.18	-0.14
69	0.97	-0.23	-0.31	-0.26	-0.19	-0.19

For each variable, the correlation coefficient of corresponding weight map with the pattern of assignation map for each cancer class can be calculated:

CC(i) = 0: indicates no correlation between biomarker and the cancer class i;

1 > CC(i) > 0: accordance between the biomarker intensity and cancer class i.

0<CC<-1: an inverse correlation between biomarker value and cancer class i.

The CC values >0.5 or <-0.5 are colored. For example, the weight map of biomarker 1282 is highly correlated with the pattern of cancer class 4 on the assignation map, and it may be an important biomarker for the samples from patients with myeloma.









18%

16%

Cationic





COMPLEMENT



C1QA C1QB C1QC C1R C1S C02 C3 C04 C4B C04B C05 C05 C06 C7 C08A C08B C08B C09 CFAB CFAD CFAD FHR1 FHR2 FHR5 FHR5 CFAD







Fig S1. Classification of identified coronas by sensor array elements according to their physiological functions, including (A) acute phase, (B) coagulation, (C) immunoglobulins, (D) lipoproteins, (E) tissue leakage, (F) complement, and (G) other plasma proteins, in human plasma of healthy subjects and patients having different types of cancers (y-axis is the percentage normalized spectral count for proteins).



Fig. S2 Schematic representation of study outline. Informative variable selection and classification model building.



Fig. S3. ROC (receiver operating characteristic) plots derived from PLS-DA and confusion matrix drived from CPANN (counter propagation artificial neural network) based on the top 69 ranked variables for five cancer and control classes. (A-F) ROC plot of sensitivity (True Positive Rate, Y-axis) versus 1 – specificity (False Positive Rate, X-axis) based on a PLS-DA built upon the 69 markers with the highest contribution for six classes. (G) Obtained Confusion matrix by CPANN indicates good prediction for all classes.



Fig. S4. Data analysis using CPANN on the data matrix corresponding to protein corona of each liposome separately. Assignation map obtained by all variables of protein corona of each liposome (anionic (A), cationic (B) and neutral (C)). As illustrated in these assignation maps, no one class of the liposomes could discriminate all 6 groups of samples as well as the composite response of the full array. The classification error for the model of anionic, cationic and neutral is 54%, 24% and 10%, respectively. Therefore, we used the protein corona profiles of all three liposomes to get the benefit of all sensor information probing cancer-specific signatures.