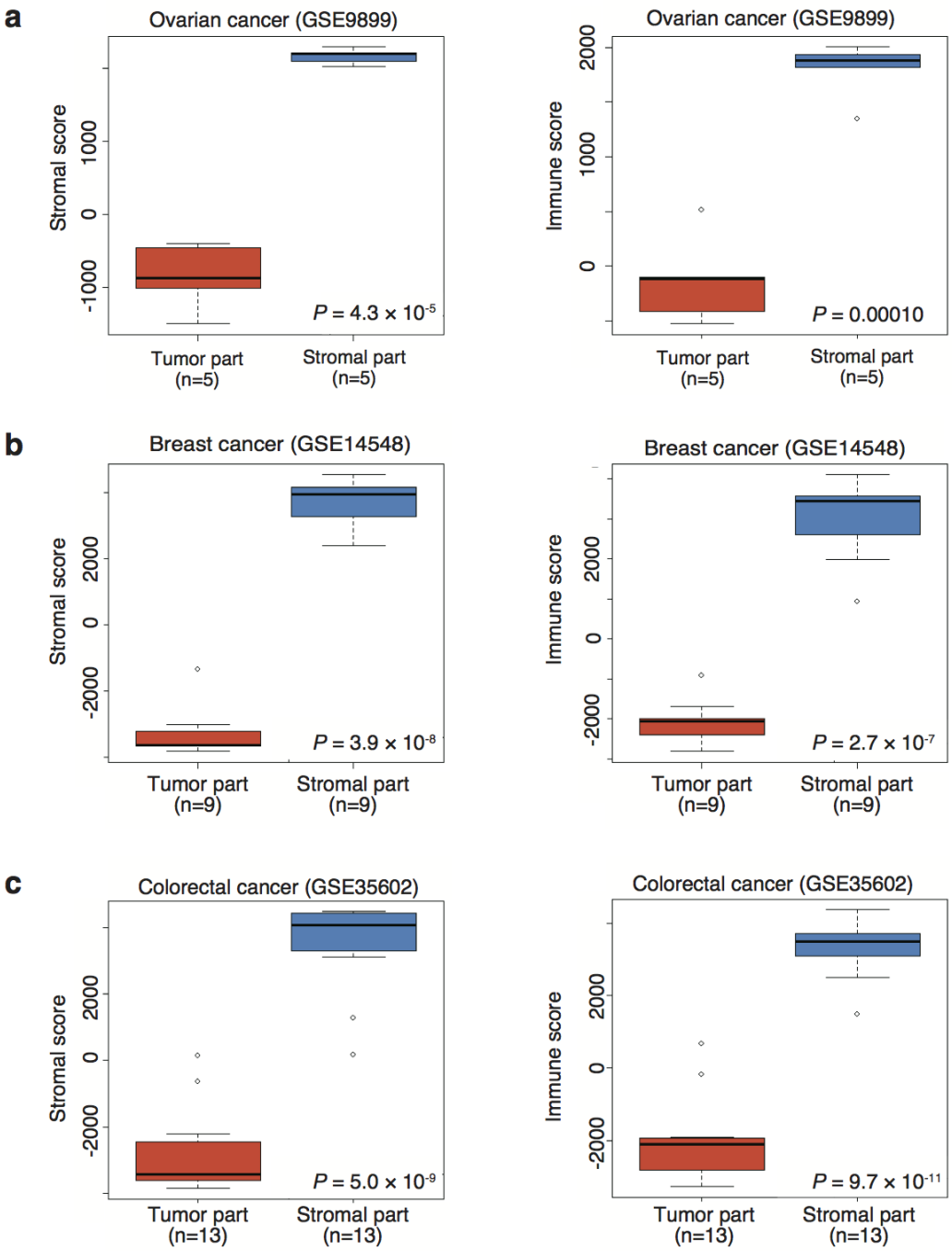
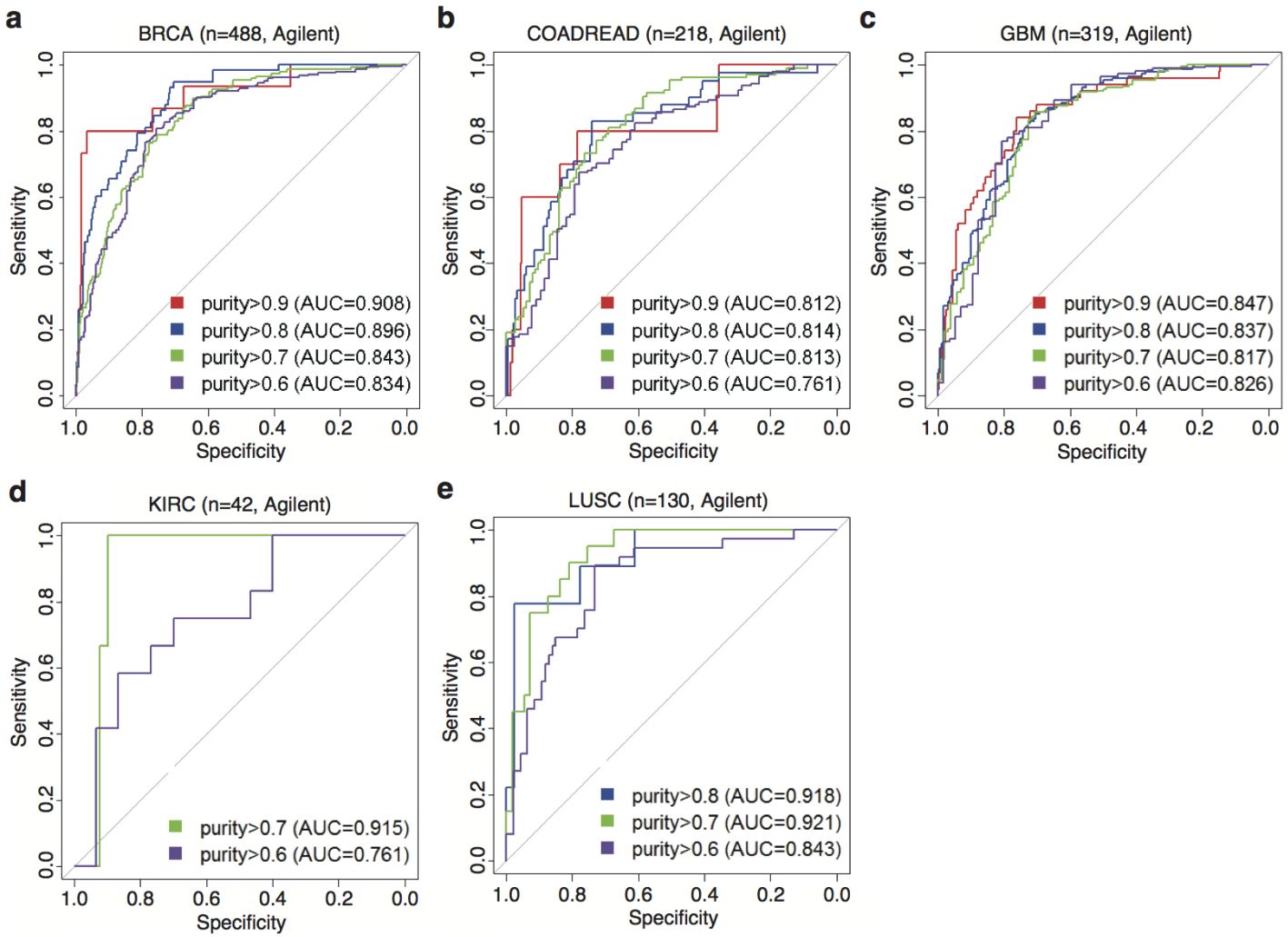
Supplementary Information

Inferring tumour purity and stromal and immune cell admixture from expression data

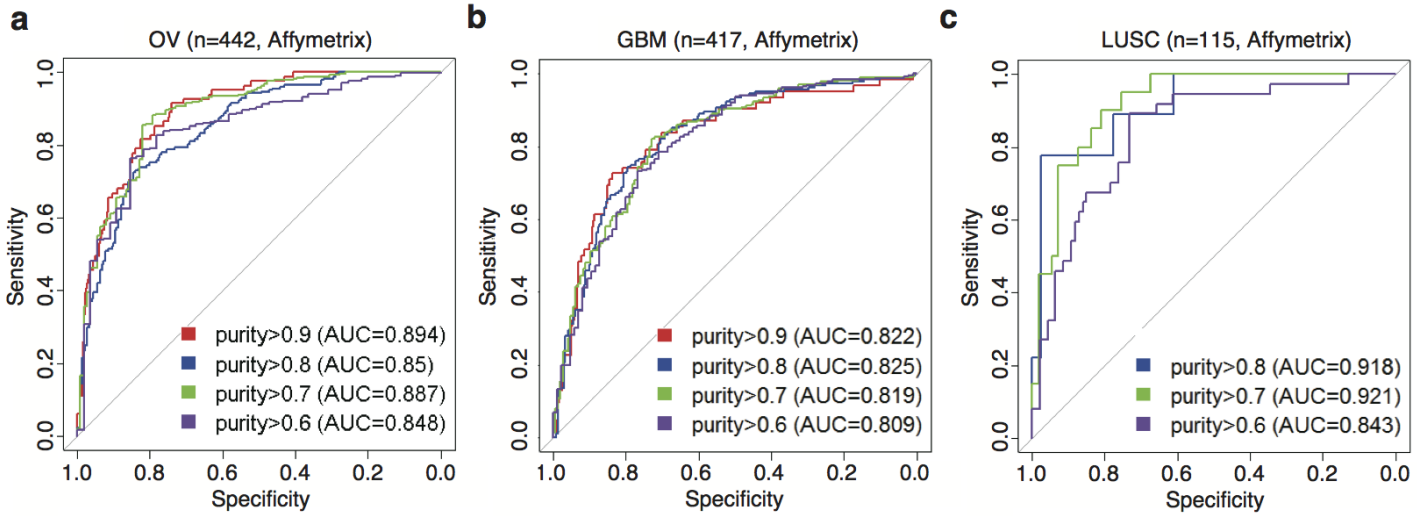


Supplementary Figure S1. Reduced stromal and immune scores in the tumor cell enriched fraction after laser capture microdissection

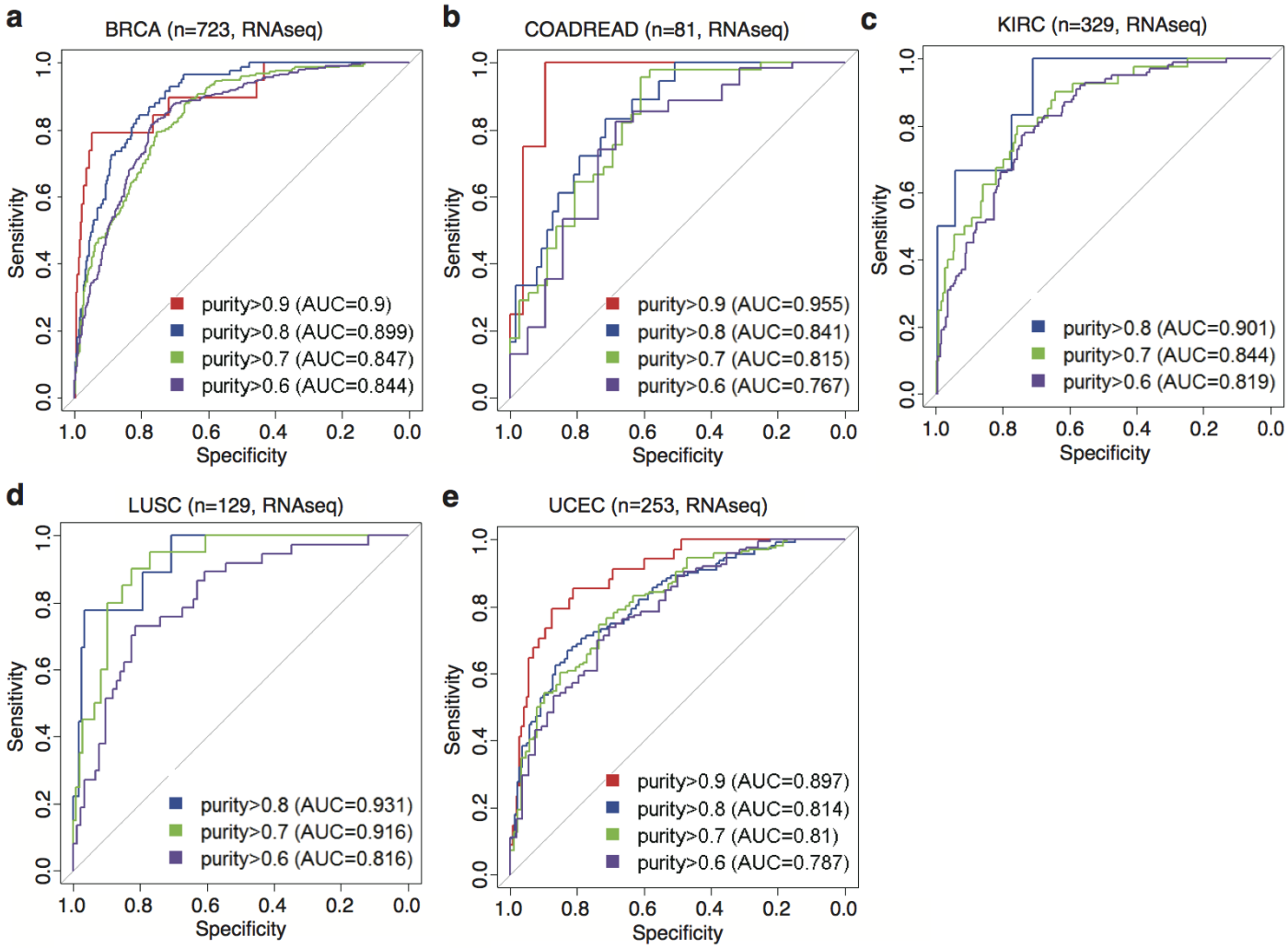
Comparison of stromal scores between tumor part and stromal part after laser capture microdissection in (a) ovarian, (b) breast, and (c) colorectal cancers. Scores of the tumor portion are shown in the left box, in red, scores of the stromal portion are shown on the right side, in blue. The box represents the interquartile range of the scores and the whiskers indicate the range of values used, within the following range: Q1 - 1.5* IQR to Q3 + 1.5*IQR where Q is a quantile and IQR is the interquartile range.



Supplementary Figure S2. Evaluation of the ESTIMATE algorithm in tumor profiled on the Agilent platform.
In Agilent platform, the accuracy of ESTIMATE algorithm was evaluated based on the area under ROC curve (AUC) per each tumor when tumor samples were divided into high and low purity groups on the basis of DNA copy number-based tumor purity.

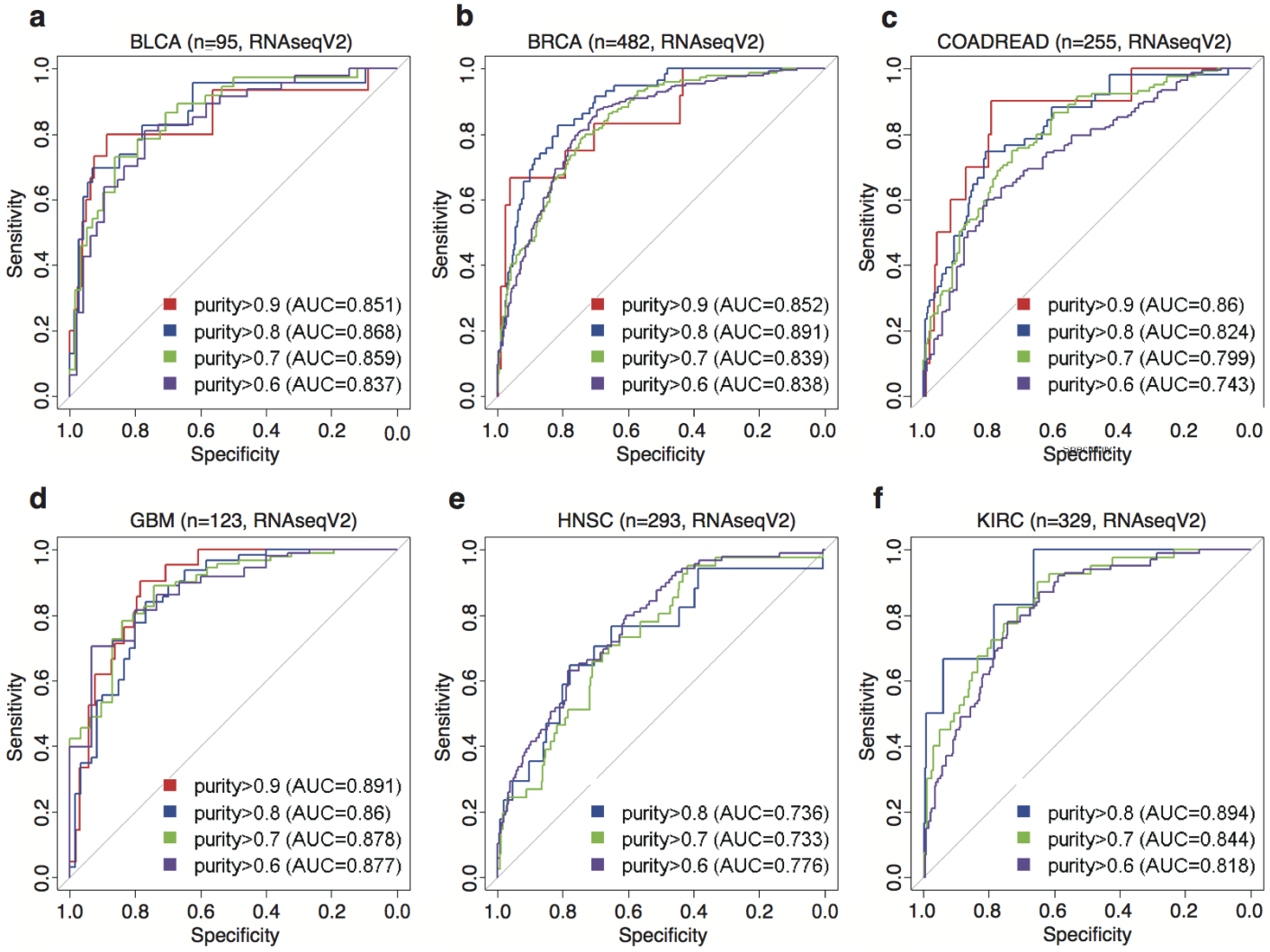


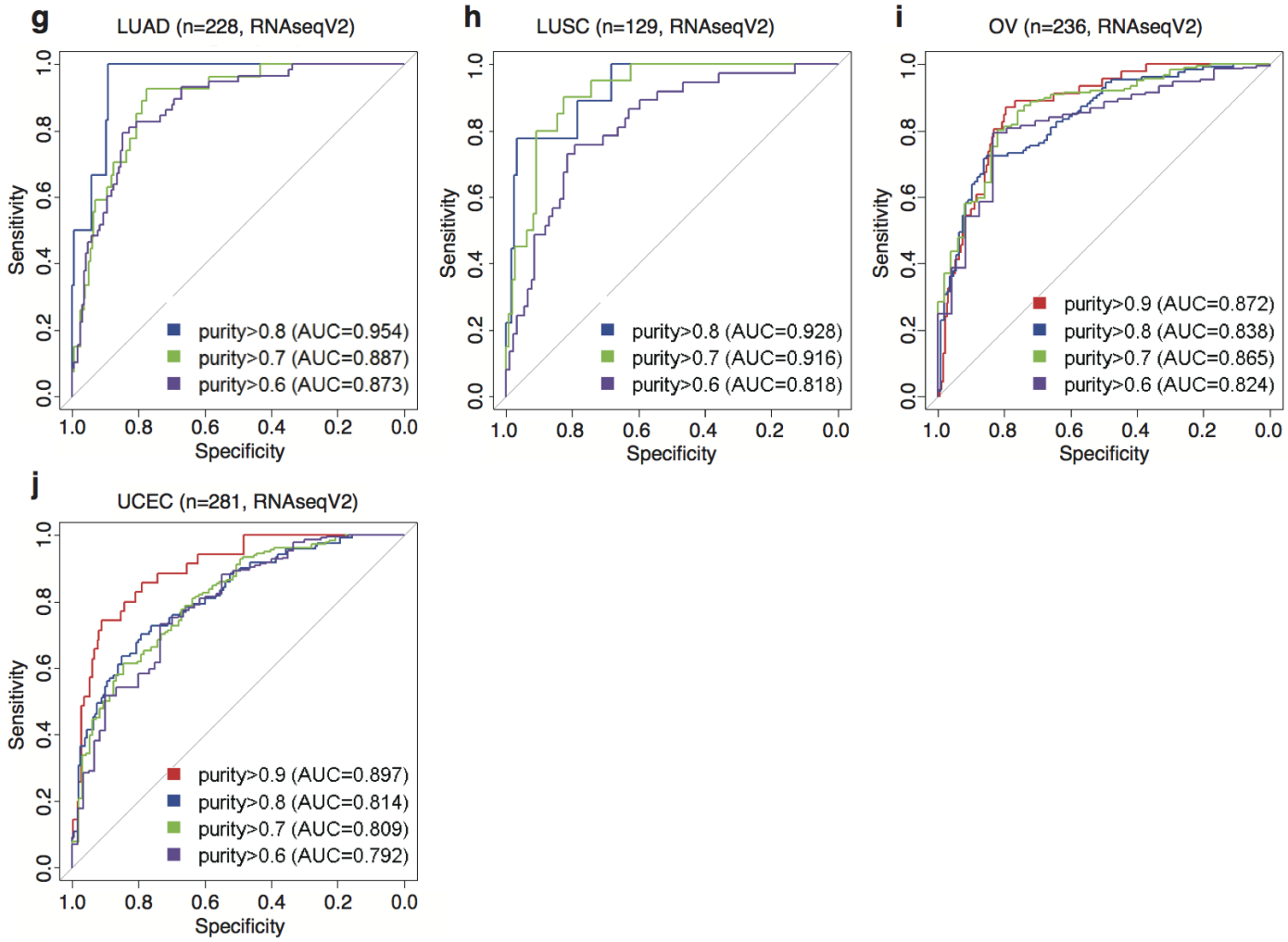
Supplementary Figure S3. Evaluation of the ESTIMATE algorithm in tumor profiled on the Affymetrix platform.
In Affymetrix platform, the accuracy of ESTIMATE algorithm was evaluated based on the area under ROC curve (AUC) per each tumor when tumor samples were divided into high and low purity groups on the basis of DNA copy number-based tumor purity. TCGA Ovarian cancer samples used in the gene selection were removed.



Supplementary Figure S4. Evaluation of the ESTIMATE algorithm in tumor profiled using RNA-sequencing and Reads Per Kilobase per Million mapped reads [RPKM]

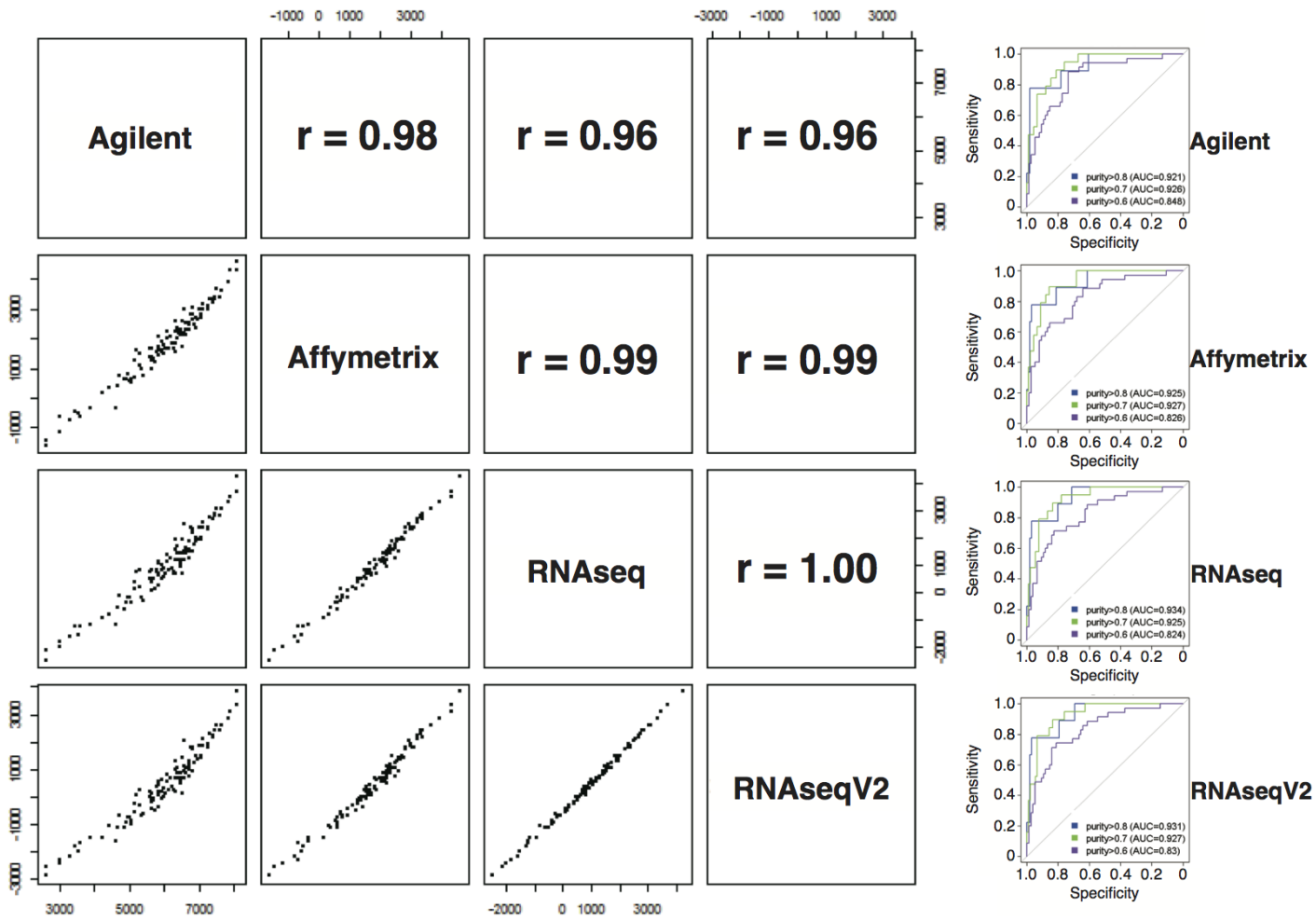
In RNA-seq platform, the accuracy of ESTIMATE algorithm was evaluated based on the area under ROC curve (AUC) per each tumor when tumor samples were divided into high and low purity groups on the basis of DNA copy number-based tumor purity.


a
BLCA (n=95, RNAseqV2)
purity>0.9 (AUC=0.851)
purity>0.8 (AUC=0.868)
purity>0.7 (AUC=0.859)
purity>0.6 (AUC=0.837)
b
BRCA (n=482, RNAseqV2)
purity>0.9 (AUC=0.852)
purity>0.8 (AUC=0.891)
purity>0.7 (AUC=0.839)
purity>0.6 (AUC=0.838)
c
COADREAD (n=255, RNAseqV2)
purity>0.9 (AUC=0.86)
purity>0.8 (AUC=0.824)
purity>0.7 (AUC=0.799)
purity>0.6 (AUC=0.743)
d
GBM (n=123, RNAseqV2)
purity>0.9 (AUC=0.891)
purity>0.8 (AUC=0.86)
purity>0.7 (AUC=0.878)
purity>0.6 (AUC=0.877)
e
HNSC (n=293, RNAseqV2)
purity>0.8 (AUC=0.736)
purity>0.7 (AUC=0.733)
purity>0.6 (AUC=0.776)
f
KIRC (n=329, RNAseqV2)
purity>0.8 (AUC=0.894)
purity>0.7 (AUC=0.844)
purity>0.6 (AUC=0.818)
Sensitivity
Specificity
0.0
0.2
0.4
0.6
0.8
1.0

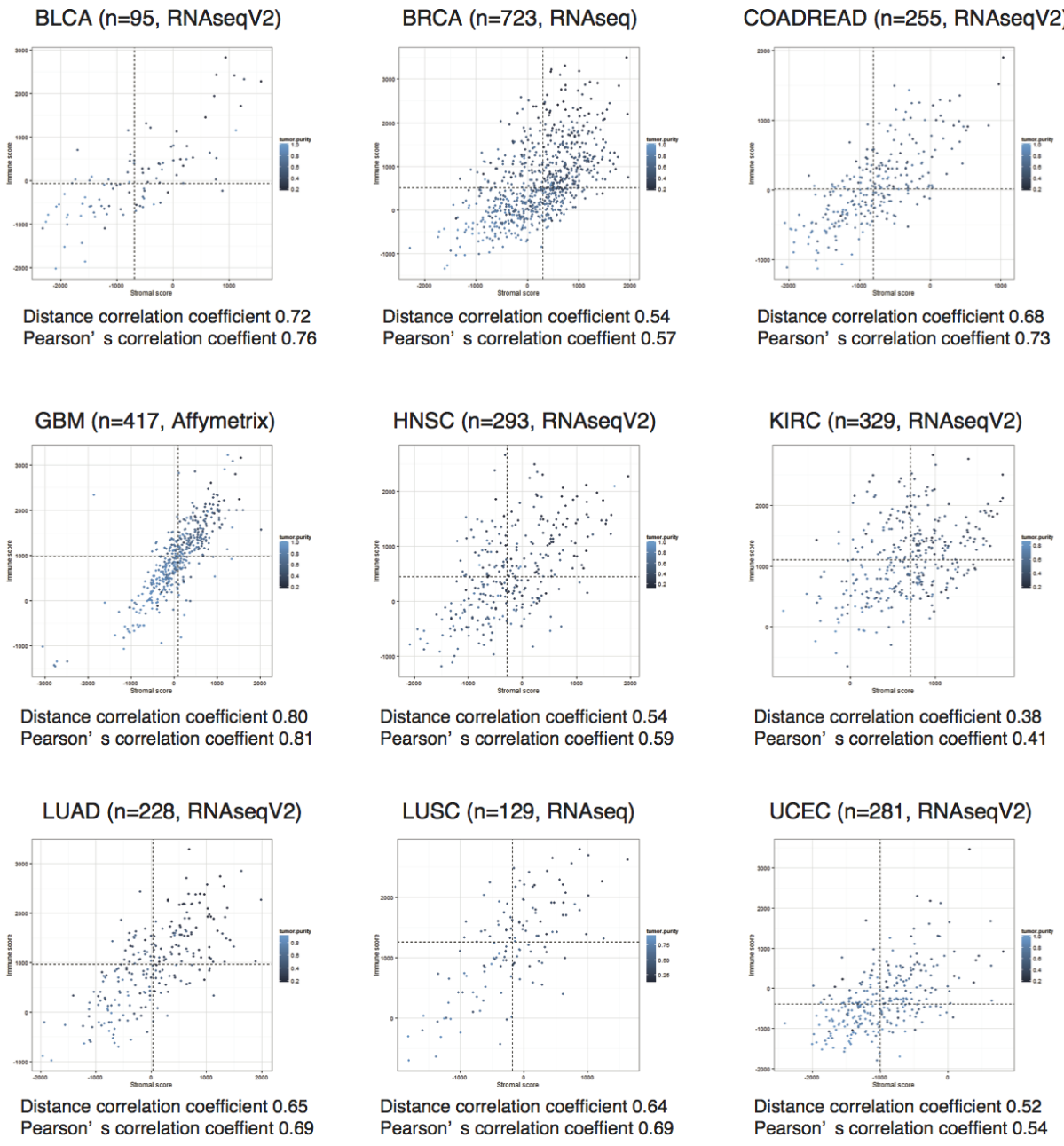


Supplementary Figure S5. Evaluation of the ESTIMATE algorithm in tumor profiled using RNA-sequencing and quantified through Expectation Maximization [RSEM]

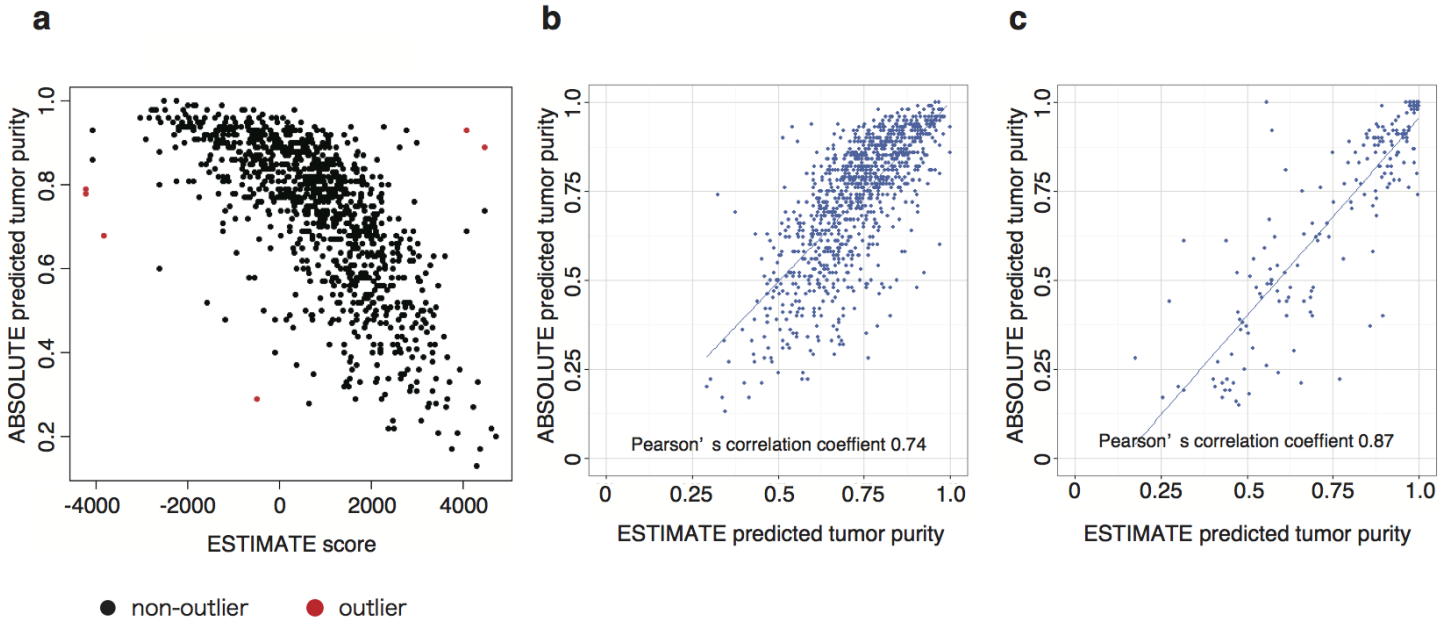
In RNA-seq platform, the accuracy of ESTIMATE algorithm was evaluated based on the area under ROC curve (AUC) per each tumor when tumor samples were divided into high and low purity groups on the basis of DNA copy number-based tumor purity. Ovarian cancer samples in the TCGA data set that were used in the gene selection were removed.



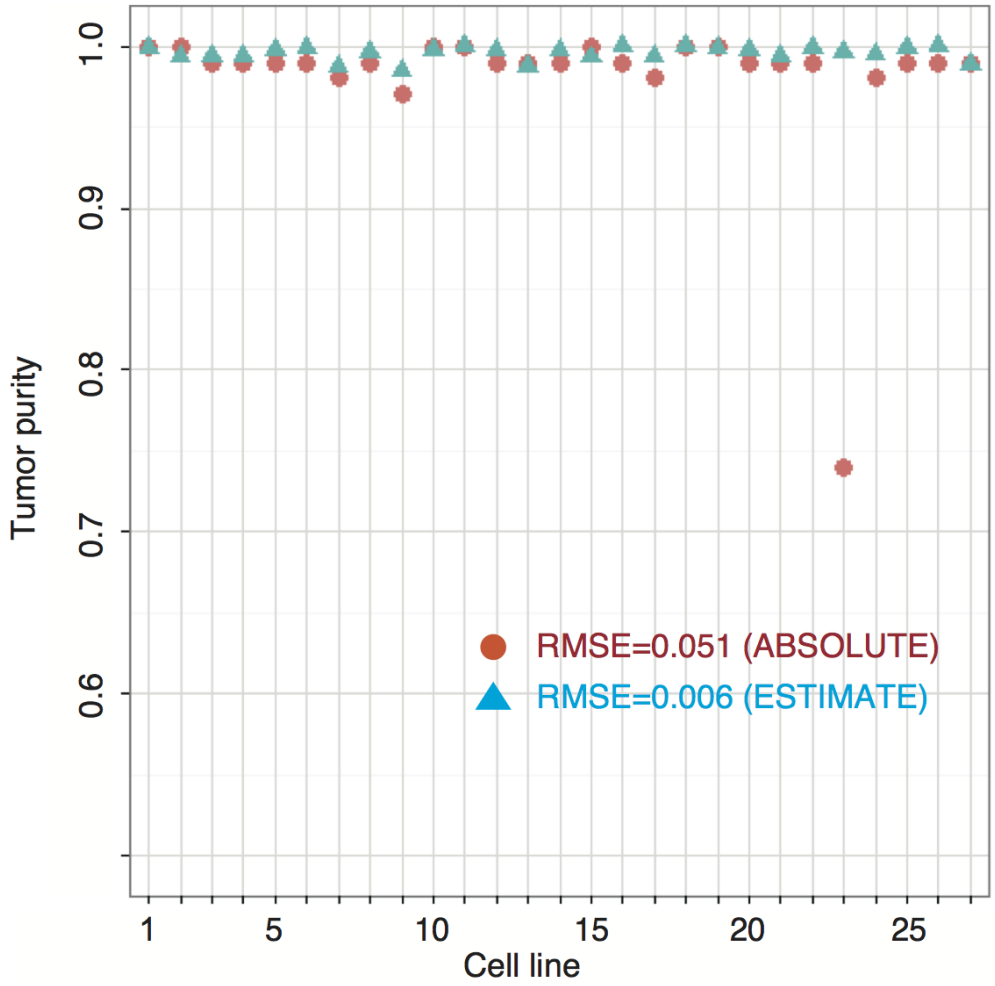
Supplementary Figure S6. Correlation between various expression platforms and ESTIMATE scores.
High correlation in ESTIMATE scores between each platform and similar AUC per platform were observed when we evaluated the prediction ability of tumor purity per each platform using 110 lung squamous cell carcinoma samples with expression data from four platforms.



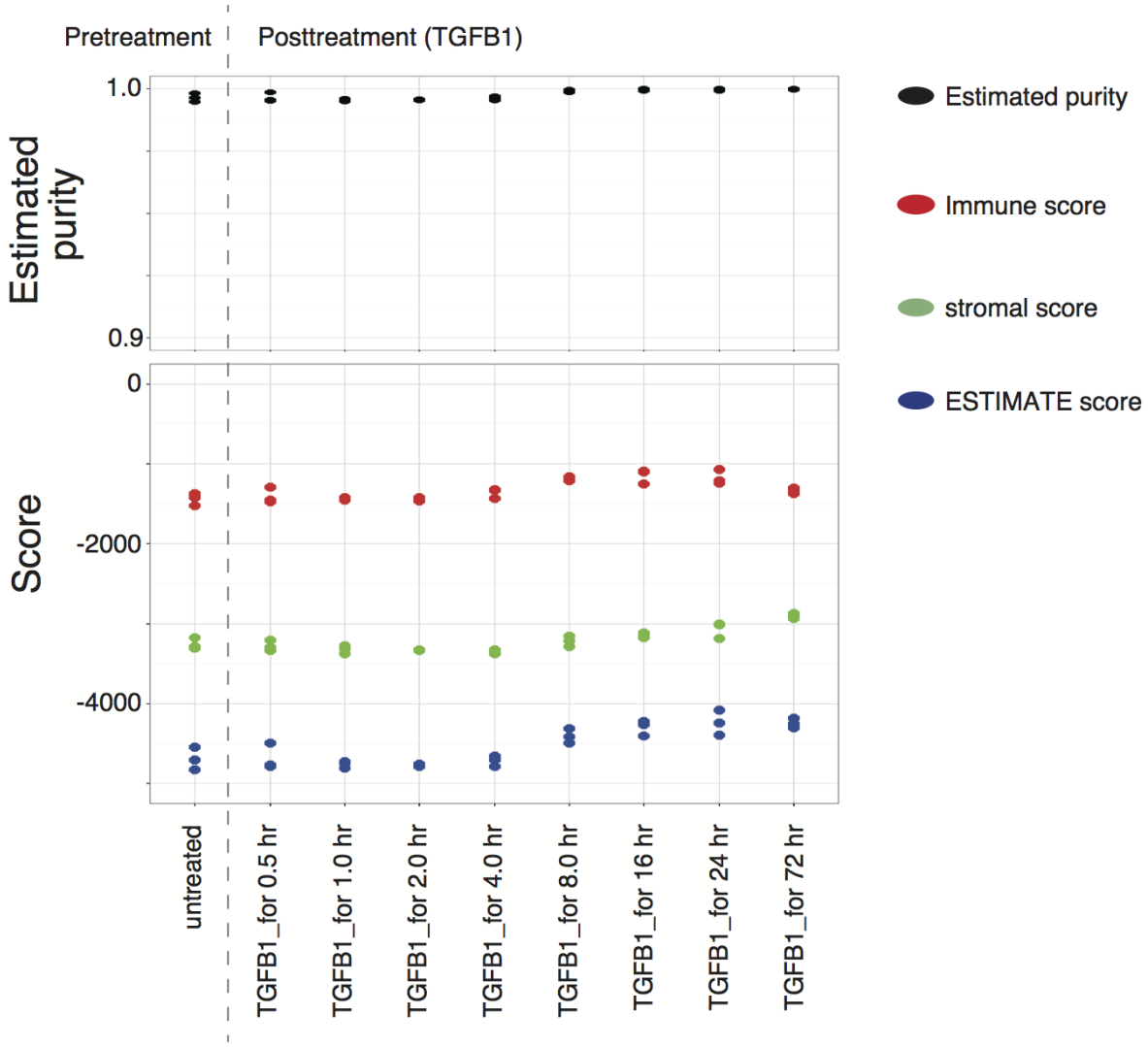
Supplementary Figure S7. The correlation between stromal and immune scores.
Scatterplots between stromal and immune scores in nine TCGA data sets. Correlation coefficient was calculated by Pearson's and distance correlation analysis.



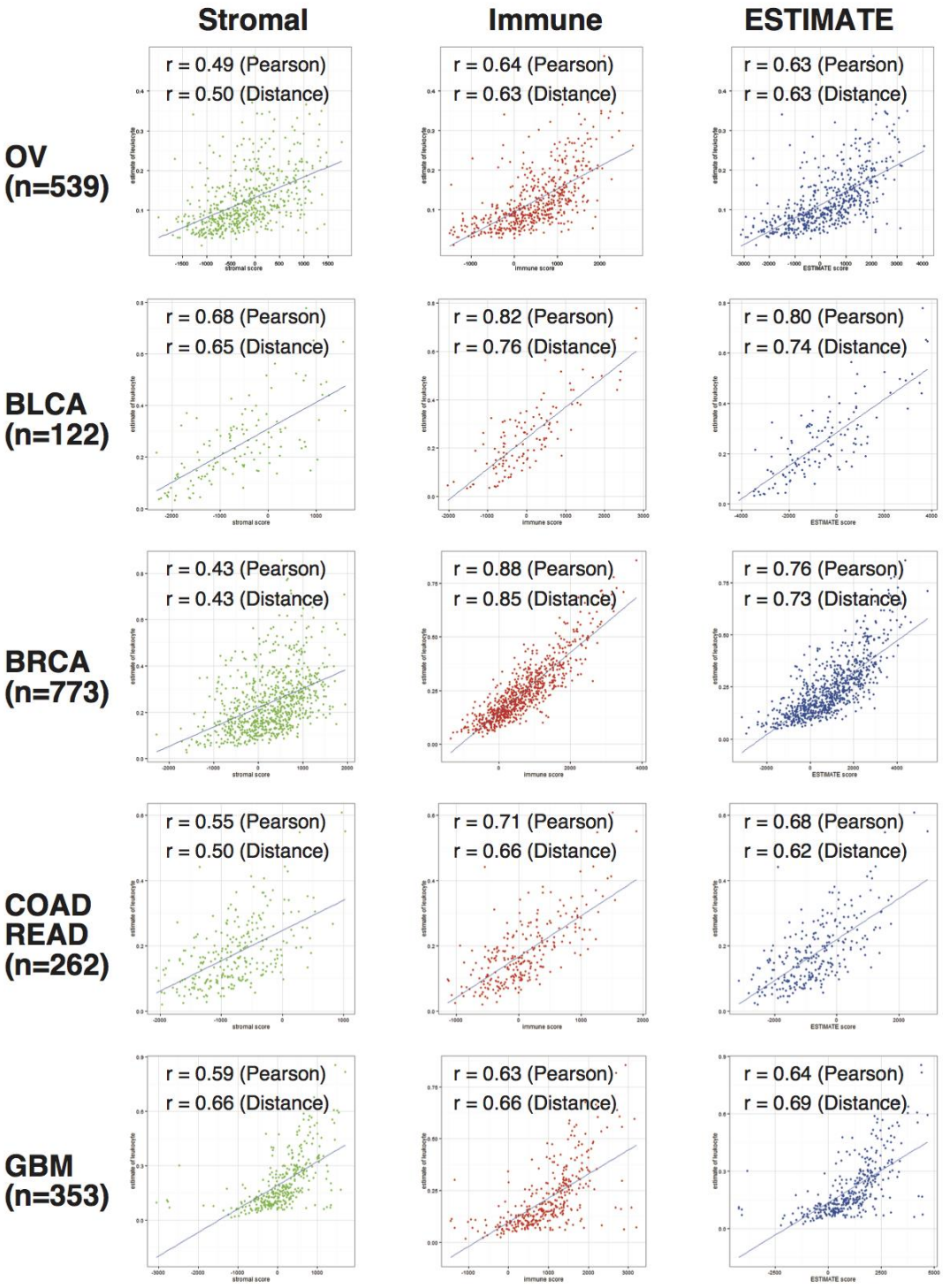
Supplementary Figure S8. The estimation of tumor purity based on ESTIMATE algorithm using Affymetrix expression data. (a) Of 1,001 TCGA Affymetrix data, six outlier samples were detected by computing a multivariate outlier criterion based on Generalized Extreme Studentized Deviate (GESD) test. (b-c) High correlation between ESTIMATE based tumor purity (x-axis) and ABSOLUTE based tumor purity (y-axis) in (b) the TCGA Affymetrix data set and (c) the independent validation data set.

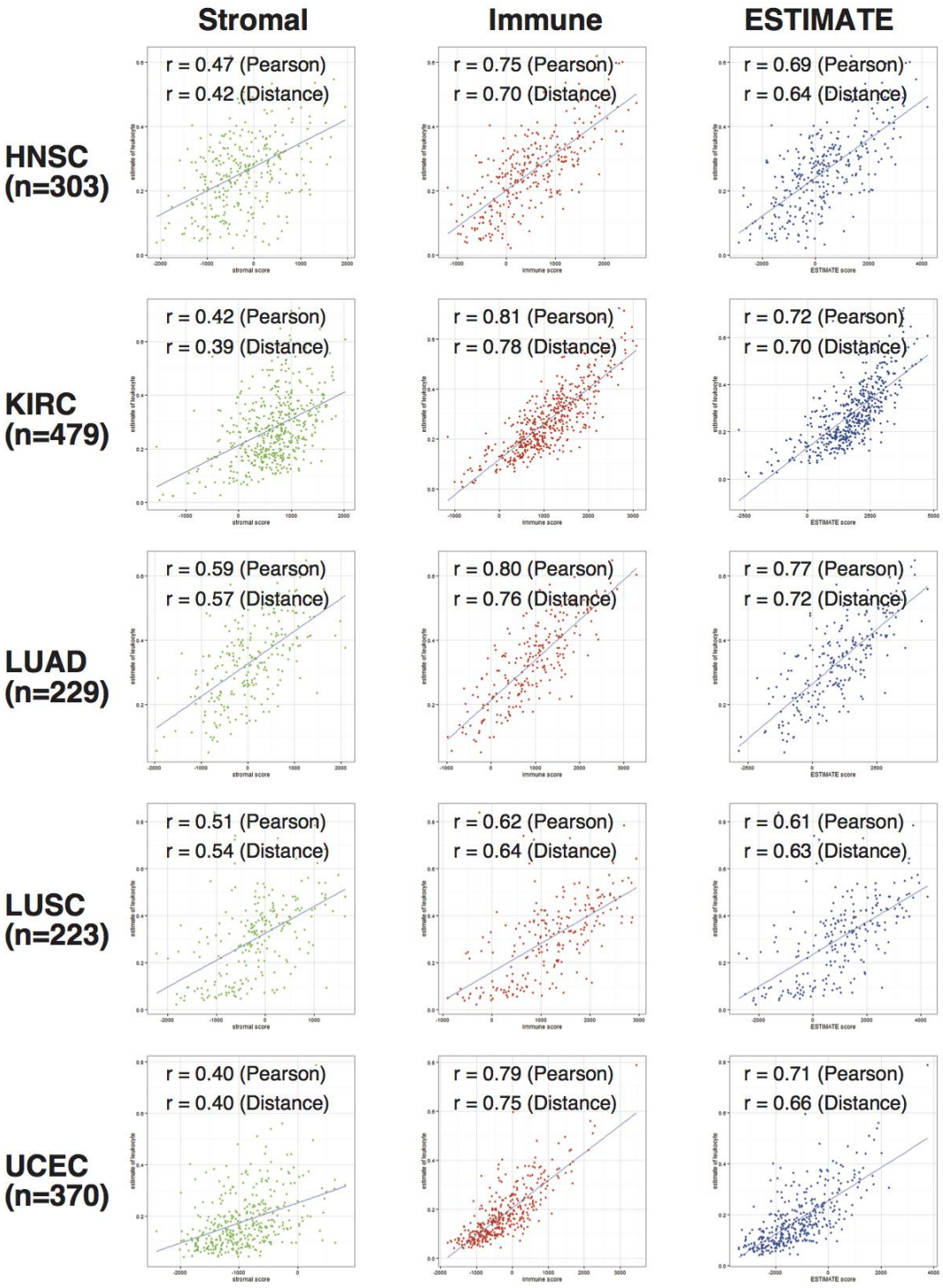


Supplementary Figure S9. Predicting tumor purity in cell lines using ESTIMATE and ABSOLUTE algorithms
A plot of tumor purity per cancer cell line sample. In twenty-seven samples (GSE34211) that had not used in the process of gene selection, both ESTIMATE and ABSOLUTE were performed. The x-axis and y-axis denote cancer cell line samples and tumor purity, respectively. RMSE stands for root mean square error.

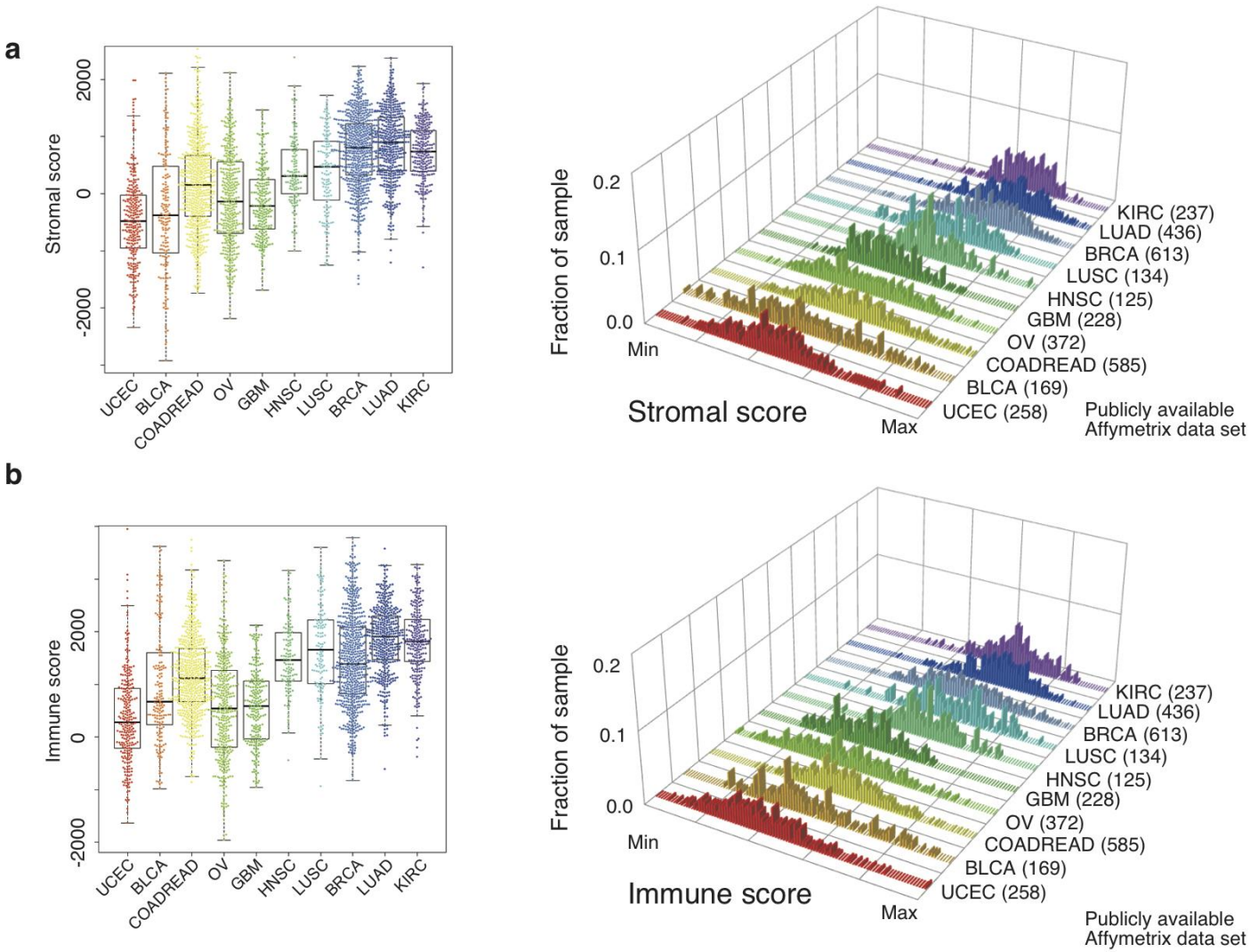


Supplementary Figure S10. ESTIMATE predictive ability is independent of altered gene expression in tumor cells
Using 26 microarray expression data from lung adenocarcinoma cell lines (GSE17708) that were treated or untreated by TGFB1 protein, stromal, immune, and ESTIMATE score were calculated. The x-axis denotes cancer cell line samples and y-axis represented mRNA expression data, estimated tumor purity, and raw stromal, immune, and ESTIMATE scores.


Stromal
Immune
ESTIMATE
OV (n=539)
r = 0.49 (Pearson)
r = 0.50 (Distance)
r = 0.64 (Pearson)
r = 0.63 (Distance)
r = 0.63 (Pearson)
r = 0.63 (Distance)
BLCA (n=122)
r = 0.68 (Pearson)
r = 0.65 (Distance)
r = 0.82 (Pearson)
r = 0.76 (Distance)
r = 0.80 (Pearson)
r = 0.74 (Distance)
BRCA (n=773)
r = 0.43 (Pearson)
r = 0.43 (Distance)
r = 0.88 (Pearson)
r = 0.85 (Distance)
r = 0.76 (Pearson)
r = 0.73 (Distance)
COAD READ (n=262)
r = 0.55 (Pearson)
r = 0.50 (Distance)
r = 0.71 (Pearson)
r = 0.66 (Distance)
r = 0.68 (Pearson)
r = 0.62 (Distance)
GBM (n=353)
r = 0.59 (Pearson)
r = 0.66 (Distance)
r = 0.63 (Pearson)
r = 0.66 (Distance)
r = 0.64 (Pearson)
r = 0.69 (Distance)

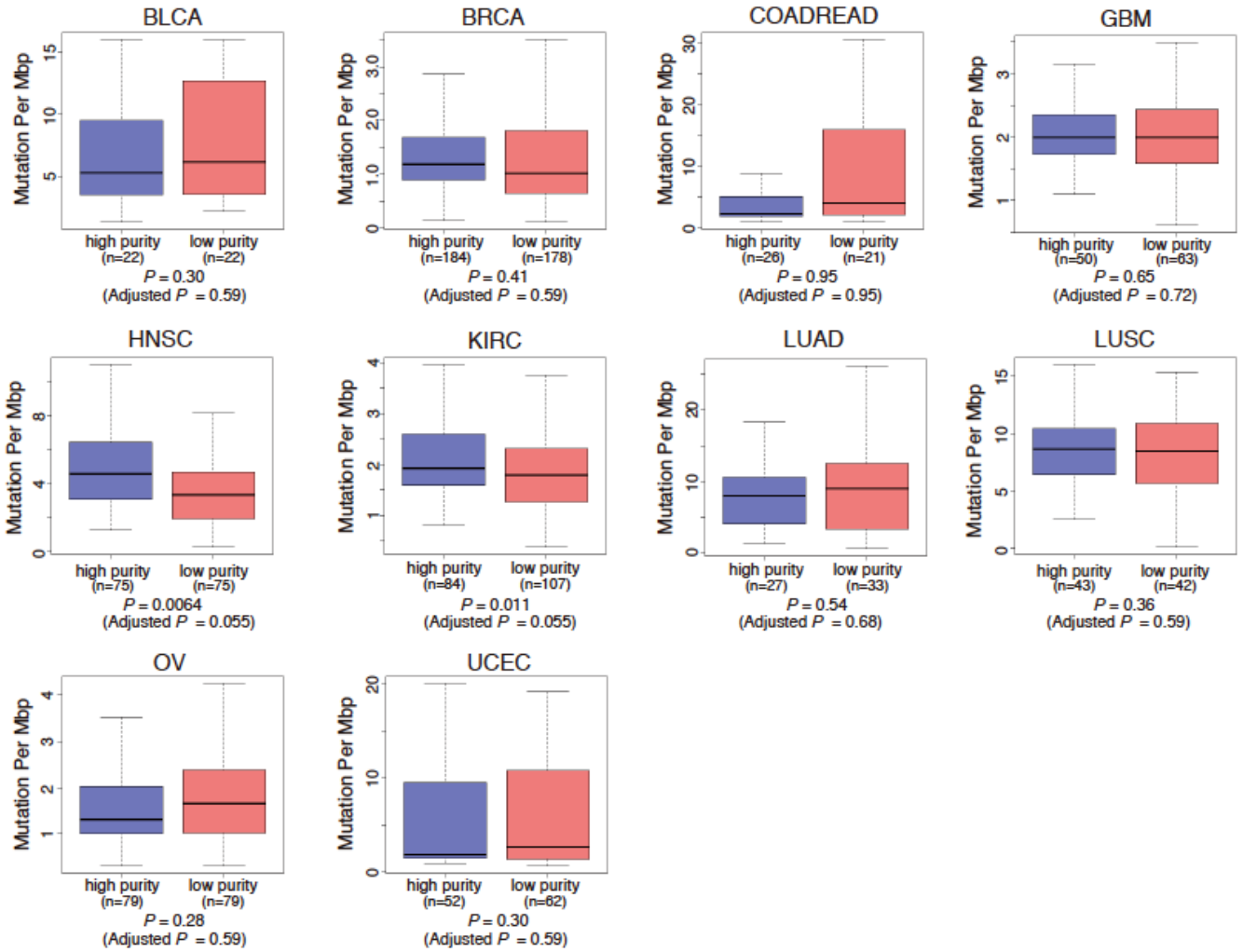


Supplementary Figure S11. High correlation between immune score and leukocyte methylation score.
Using 3,681 samples with leukocyte methylation score, correlation analysis between stromal (green), immune (red), and ESTIMATE score (blue) and leukocyte methylation score was performed in the discovery set (OV) and the validation data sets. TCGA ovarian cancer samples used in the gene selection were removed.



Supplementary Figure S12. Beeswarm plots of ESTIMATE scores in non-TCGA data

Distinct distributions of ESTIMATE scores across different tumor types were observed in Affymetrix platform data sets. The one-way Analysis of Variance (ANOVA) followed by Tukey's post hoc test was performed to examine the differences in these scores among different tumor types. P-values for Tukey's post hoc test are listed in Supplementary Table S4. The boxes in the boxplots in figure A and B represent the interquartile range of the scores and the whiskers indicate the range of values used, within the following range: Q1 - 1.5* IQR to Q3 + 1.5*IQR where Q is a quantile and IQR is the interquartile range.



Supplementary Figure S13. The relationship between the level of tumor purity based on ESTIMATE and mutation frequency.
We compared differences in mutation frequency between high and low purity subgroups by using the unpaired t-test. . The box represents the interquartile range of the scores and the whiskers indicate the range of values used, within the following range: Q1 - 1.5* IQR to Q3 + 1.5*IQR where Q is a quantile and IQR is the interquartile range.

Supplementary Table S1. A list of public data sets used in this study

Publicly available microarray expression and SNP array copy number data sets used in this study are listed in this table.

Dataset	Type	Platform	Disease	Reference
Detection of stromal/immune cell related genes				
GSE1133	EXP	Affymetrix HG-U133A	Normal tissue	[70]
GSE9899 (GSE9890/9891)	EXP*	Affymetrix HG-U133Plus2.0	OV	[71]
GSE14548	EXP*	Affymetrix Human X3P	BRCA	[72]
GSE35602	EXP*	Agilent 4x44K (G4112F)	COADREAD	[73]
Cancer Cell line Encyclopedia	EXP	Affymetrix HG-U133Plus2.0	Cell line	[74]
Glioma stem cell line data	EXP	Affymetrix HG-U133A2	Cell line	[75, 76, 77, 78, 79, 80, 81]
Validation of stromal/immune signatures				
GSE29156	EXP*	Affymetrix Human Exon1.0 ST	OV	-
GSE10797	EXP*	Affymetrix HG-U133A2	BRCA	[82]
GSE33363	EXP*	Affymetrix HG-U133Plus2.0	LUAD	[83]
GSE17708	EXP	Affymetrix HG-U133Plus2.0	Cell line	[84]
GSE18520	EXP	Affymetrix HG-U133Plus2.0	Normal ovarian epithelium cell	[85]
Validation of ESTIMATE				
GSE14994	EXP/SNP	Affymetrix HT-HG-U133A & Sty250K	KIRC	[86]
GSE19949	EXP/SNP	Affymetrix HT-HG-U133A & SNP6	KIRC	[87]
GSE36895/GSE25540	EXP/SNP	Affymetrix HG-U133Plus2.0 & SNP6	KIRC	[88]
GSE27854/GSE27910	EXP/SNP	Affymetrix HG-U133Plus2.0 & Sty250K	COADREAD	[89]
GSE33356	EXP/SNP	Affymetrix HG-U133Plus2.0 & SNP6	Lung	[90]
GSE34211	EXP/SNP	Affymetrix HG-U133Plus2.0 & SNP6	Cell line (8 breast, 4 brain, 1 colon, 5 lung, 3 ovary, 5 renal, and 1 melanoma)	[91]

*The expression data were obtained from tumor-cell enrich samples after laser-capture microdissection.

EXP: expression, SNP: Single Nucleotide Polymorphism

Supplementary Table S1. A list of public data sets used in this study (Continued)

Publicly available microarray expression and SNP array copy number data sets used in this study are listed in this table.

Dataset	Type	Platform	Disease	Reference
Investigating distribution of stromal/immune score				
GSE2109	EXP	Affymetrix HG-U133Plus2.0	BLCA, BRCA, COADREAD, KIRC, LUAD, LUSC, OV, UCEC	-
GSE3292	EXP	Affymetrix HG-U133Plus2.0	HNSC	[92]
GSE9891	EXP	Affymetrix HG-U133Plus2.0	OV	[71]
GSE17025	EXP	Affymetrix HG-U133Plus2.0	UCEC	[93]
GSE17538	EXP	Affymetrix HG-U133Plus2.0	COADREAD	[94]
GSE20685	EXP	Affymetrix HG-U133Plus2.0	BRCA	[95]
GSE28571	EXP	Affymetrix HG-U133Plus2.0	LUAD, LUSC	[96]
GSE31189	EXP	Affymetrix HG-U133Plus2.0	BLCA	[97]
GSE31684	EXP	Affymetrix HG-U133Plus2.0	BLCA	[98]
GSE31210	EXP	Affymetrix HG-U133Plus2.0	LUAD	[99]
GSE36895	EXP	Affymetrix HG-U133Plus2.0	KIRC	[88]
GSE37745	EXP	Affymetrix HG-U133Plus2.0	LUAD, LUSC	[100]
E-MTAB-1328	EXP	Affymetrix HG-U133Plus2.0	HNSC	[101]
Rembrandt data†	EXP	Affymetrix HG-U133Plus2.0	GBM	[102]

† The Rembrandt data were downloaded from National Cancer Institute web page (https://caintegrator.nci.nih.gov/rembrandt/).

EXP: expression

Supplementary Table S2. The accuracy of ESTIMATE-based tumor purity in the validation data set.

ESTIMATE algorithm was validation in the independent data set consisting of six publicly available data sets. The sensitivity, specificity, and accuracy per cutoff was calculated by pROC.

Cutoff	Sensitivity	Specificity	Accuracy
0.9	0.820	0.855	0.831
0.8	0.763	0.939	0.851
0.7	0.840	0.967	0.918
0.6	0.694	0.955	0.872

Supplementary Table S3. Stromal, Immune, and ESTIMATE score in ten normal ovarian epithelium samples

GEO_ID	Tissue	Stromal score	Immune score	ESTIMATE score	ESTIMATE-predicted tumor purity
GSM462643	normal	912.0	2279.9	3191.9	0.477
GSM462644	normal	191.5	396.4	587.9	0.770
GSM462645	normal	142.5	902.6	1045.1	0.726
GSM462646	normal	-136.8	49.8	-87.1	0.830
GSM462647	normal	650.4	2330.9	2981.3	0.504
GSM462648	normal	350.4	1058.8	1409.2	0.688
GSM462649	normal	585.5	1831.7	2417.1	0.574
GSM462650	normal	325.9	764.6	1090.5	0.721
GSM462651	normal	260.7	744.6	1005.3	0.730
GSM462652	normal	114.6	202.7	317.3	0.795

Supplementary Table S4. Tukey's post hoc test between two different tumor types.

The adjusted *P*-values for Tukey's post hoc test between two different tumor types are listed in this table.

TCGA (RNAseqV2)	Stromal score *P*-value	Immune score *P*-value	ESTIMATE score *P*-value
BRCA-BLCA	0	0.83	0
COADREAD-BLCA	0.58	1.0	1.0
GBM-BLCA	0.00063	0.26	0.0077
HNSC-BLCA	6.7E-08	1.1E-07	0
KIRC-BLCA	0	0	0
LUAD-BLCA	0	0	0
LUSC-BLCA	2.6E-08	8.1E-10	0
OV-BLCA	1.0	0.50	1.0
UCEC-BLCA	0.0013	0.053	0.0025
COADREAD-BRCA	0	0.95	0
GBM-BRCA	0	0.88	0.00084
HNSC-BRCA	0	8.6E-10	0.75
KIRC-BRCA	0	0	0
LUAD-BRCA	4.5E-06	0	2.4E-05
LUSC-BRCA	0	0	1.0
OV-BRCA	0	7.1E-06	0
UCEC-BRCA	0	0	0
GBM-COADREAD	0	0.34	9.7E-06
HNSC-COADREAD	0	0	0
KIRC-COADREAD	0	0	0
LUAD-COADREAD	0	0	0
LUSC-COADREAD	0	0	0
OV-COADREAD	0.013	0.022	1.0
UCEC-COADREAD	0.20	5.1E-05	0.00088
HNSC-GBM	0.96	0.025	0.19
KIRC-GBM	0	0	0
LUAD-GBM	5.21E-05	0	0
LUSC-GBM	0.81	0.0013	0.029
OV-GBM	0.00075	5.2E-06	3.8E-06
UCEC-GBM	0	4.4E-10	0
KIRC-HNSC	0	0	0
LUAD-HNSC	0.00053	1.0E-08	8.6E-08
LUSC-HNSC	1.0	0.98	0.99
OV-HNSC	0	0	0
UCEC-HNSC	0	0	0
LUAD-KIRC	0	0.30	0
LUSC-KIRC	0	0	0
OV-KIRC	0	0	0
UCEC-KIRC	0	0	0
LUSC-LUAD	0.016	7.5E-05	0.00014
OV-LUAD	0	0	0
UCEC-LUAD	0	0	0
OV-LUSC	0	0	0
UCEC-LUSC	0	0	0
UCEC-OV	2.67E-09	0.98	0.0024

Supplementary Table S4. Tukey's post hoc test between two different tumor types (Continued)
The adjusted *P*-values for Tukey's post hoc test between two different tumor types are listed in this table.

Publicly available data (Affymetrix)	Stromal score *P*-value	Immune score *P*-value	ESTIMATE score *P*-value
BRCA-BLCA	0	0	0
COADREAD-BLCA	0	0.054	5.0E-07
KIRC-BLCA	0	0	0
LUAD-BLCA	0	0	0
LUSC-BLCA	0	0	0
OV-BLCA	0.037	3.0E-05	0.97
UCEC-BLCA	0.71	0	1.5E-05
GBM-BLCA	0.52	0.00026	0.87
HNSC-BLCA	0	5.8E-07	0
COADREAD-BRCA	0	2.0E-07	0
KIRC-BRCA	1.0	4.3E-07	0.053
LUAD-BRCA	0.55	0	1.4E-06
LUSC-BRCA	2.4E-06	0.25	0.95
OV-BRCA	0	0	0
UCEC-BRCA	0	0	0
GBM-BRCA	0	0	0
HNSC-BRCA	8.3E-06	1.0	0.36
KIRC-COADREAD	0	0	0
LUAD-COADREAD	0	0	0
LUSC-COADREAD	0.057	6.0E-08	1.2E-05
OV-COADREAD	3.2E-06	0	0
UCEC-COADREAD	0	0	0
GBM-COADREAD	5.1E-07	0	0
HNSC-COADREAD	0.060	0.0022	0.0031
LUAD-KIRC	0.61	1.0	0.95
LUSC-KIRC	0.00027	0.72	0.024
OV-KIRC	0	0	0
UCEC-KIRC	0	0	0
GBM-KIRC	0	0	0
HNSC-KIRC	0.00056	0.020	0.00089
LUSC-LUAD	6.0E-10	0.31	7.2E-05
OV-LUAD	0	0	0
UCEC-LUAD	0	0	0
GBM-LUAD	0	0	0
HNSC-LUAD	1.1E-08	0.0013	6.1E-07
OV-LUSC	1.8E-09	0	0
UCEC-LUSC	0	0	0
GBM-LUSC	0	0	0
HNSC-LUSC	1.0	0.90	1.0
UCEC-OV	1.1E-07	0.077	6.0E-05
GBM-OV	0.99	1.0	1.0
HNSC-OV	4.9E-09	0	0
GBM-UCEC	0.00060	0.17	0.0052
HNSC-UCEC	0	0	0
HNSC-GBM	0	0	0

Supplementary Table S5. Comparison of substitution frequency between high and low purity subgroups.
The adjusted *P*-value (Benjamini-Hochberg) per substitution pattern for unpaired t-test between two high and low purity subgroups are listed in this table.

			C>A		C>G		C>T		T>A		T>C		T>G	
Tumor type	Purity	No.	mean±sd	Adjusted p-value	mean±sd	Adjusted p-value	mean±sd	Adjusted p-value	mean±sd	Adjusted p-value	mean±sd	Adjusted p-value	mean±sd	Adjusted p-value
BLCA	high	22	0.12±0.071	0.8	0.25±0.10	0.8	0.48±0.074	0.8	0.030±0.024	0.8	0.084±0.049	0.97	0.030±0.024	0.8
	low	22	0.10±0.076		0.27±0.089		0.50±0.078		0.027±0.024		0.085±0.066		0.022±0.025	
BRCA	high	184	0.15±0.075	0.22	0.13±0.076	0.072	0.49±0.12	0.81	0.062±0.045	0.26	0.12±0.071	0.5	0.052±0.042	0.81
	low	178	0.13±0.086		0.15±0.11		0.49±0.13		0.054±0.051		0.11±0.077		0.054±0.069	
COAD READ	high	26	0.15±0.086	0.24	0.049±0.034	0.89	0.59±0.13	0.91	0.041±0.033	0.94	0.12±0.085	0.24	0.051±0.031	0.24
	low	21	0.12±0.048		0.043±0.040		0.60±0.11		0.042±0.029		0.16±0.089		0.037±0.021	
GBM	high	50	0.10±0.46	0.82	0.080±0.039	0.59	0.60±0.12	0.45	0.052±0.034	0.32	0.13±0.054	0.44	0.030±0.022	0.44
	low	63	0.10±0.45		0.075±0.041		0.62±0.096		0.041±0.024		0.12±0.067		0.036±0.024	
HNSC	high	75	0.18±0.090	**0.00009**	0.15±0.072	0.3	0.45±0.12	**0.015**	0.068±0.036	**0.015**	0.12±0.044	0.9	0.037±0.024	0.61
	low	75	0.11±0.067		0.17±0.096		0.51±0.13		0.052±0.036		0.12±0.088		0.041±0.045	
KIRC	high	84	0.18±0.054	0.2	0.12±0.034	0.13	0.34±0.70	0.51	0.11±0.041	0.15	0.17±0.062	0.23	0.081±0.033	0.23
	low	107	0.20±0.065		0.11±0.051		0.33±0.076		0.096±0.040		0.18±0.059		0.088±0.037	
LUAD	high	27	0.29±0.16	0.97	0.14±0.10	0.97	0.34±0.14	0.97	0.068±0.031	0.6	0.11±0.060	0.97	0.049±0.075	0.97
	low	33	0.30±0.16		0.14±0.069		0.34±0.15		0.082±0.033		0.98±0.47		0.040±0.020	
LUSC	high	43	0.30±0.080	0.24	0.15±0.051	0.36	0.31±0.079	0.49	0.090±0.026	**0.015**	0.11±0.031	0.13	0.034±0.014	0.13
	low	41	0.34±0.16		0.16±0.088		0.30±0.14		0.070±0.032		0.10±0.037		0.028±0.013	
OV	high	79	0.18±0.071	0.63	0.15±0.068	0.056	0.42±0.13	0.33	0.080±0.043	0.7	0.12±0.048	0.63	0.057±0.036	0.46
	low	79	0.17±0.072		0.17±0.069		0.38±0.12		0.077±0.042		0.13±0.059		0.064±0.034	
UCEC	high	52	0.15±0.061	0.93	0.080±0.064	0.93	0.57±0.12	0.93	0.032±0.023	0.93	0.12±0.082	0.93	0.044±0.036	0.93
	low	62	0.15±0.082		0.069±0.076		0.59±0.12		0.030±0.029		0.12±0.053		0.041±0.036	

Supplementary References

70. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D*, et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(16)**:** 6062-6067.

71. Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S*, et al.* Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2008, **14**(16)**:** 5198-5208.

72. Ma XJ, Dahiya S, Richardson E, Erlander M, Sgroi DC. Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast cancer research : BCR* 2009, **11**(1)**:** R7.

73. Nishida N, Nagahara M, Sato T, Mimori K, Sudo T, Tanaka F*, et al.* Microarray analysis of colorectal cancer stromal tissue reveals upregulation of two oncogenic miRNA clusters. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2012, **18**(11)**:** 3054-3070.

74. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S*, et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012, **483**(7391)**:** 603-607.

75. Jiang H, Gomez-Manzano C, Aoki H, Alonso MM, Kondo S, McCormick F*, et al.* Examination of the therapeutic potential of Delta-24-RGD in brain tumor stem cells: role of autophagic cell death. *Journal of the National Cancer Institute* 2007, **99**(18)**:** 1410-1414.

76. Liu D, Martin V, Fueyo J, Lee OH, Xu J, Cortes-Santiago N*, et al.* Tie2/TEK modulates the interaction of glioma and brain tumor stem cells with endothelial cells and promotes an invasive phenotype. *Oncotarget* 2010, **1**(8)**:** 700-709.

77. Wei J, Barr J, Kong LY, Wang Y, Wu A, Sharma AK*, et al.* Glioblastoma cancer-initiating cells inhibit T-cell proliferation and effector responses by the signal transducers and activators of transcription 3 pathway. *Molecular cancer therapeutics* 2010, **9**(1)**:** 67-78.

78. Bhat KP, Salazar KL, Balasubramaniyan V, Wani K, Heathcock L, Hollingsworth F*, et al.* The transcriptional coactivator TAZ regulates mesenchymal differentiation in malignant glioma. *Genes & development* 2011, **25**(24)**:** 2594-2609.

79. Zhang N, Wei P, Gong A, Chiu WT, Lee HT, Colman H*, et al.* FoxM1 promotes beta-catenin nuclear localization and controls Wnt target-gene expression and glioma tumorigenesis. *Cancer cell* 2011, **20**(4)**:** 427-442.

80. Kamal MM, Sathyan P, Singh SK, Zinn PO, Marisetty AL, Liang S*, et al.* REST regulates oncogenic properties of glioblastoma stem cells. *Stem cells* 2012, **30**(3)**:** 405-414.

81. Fu J, Koul D, Yao J, Wang S, Yuan Y, Colman H*, et al.* Novel HSP90 inhibitor NVP-HSP990 targets cell cycle regulators to ablate Olig2-positive glioma tumor initiating cells. *Cancer research* 2013.

82. Casey T, Bond J, Tighe S, Hunter T, Lintault L, Patel O*, et al.* Molecular signatures suggest a

major role for stromal cells in development of invasive breast cancer. *Breast cancer research and treatment* 2009, **114**(1)**:** 47-62.

83. Edlund K, Lindskog C, Saito A, Berglund A, Ponten F, Goransson-Kultima H*, et al.* CD99 is a novel prognostic stromal marker in non-small cell lung cancer. *International journal of cancer Journal international du cancer* 2012, **131**(10)**:** 2264-2273.

84. Sartor MA, Mahavisno V, Keshamouni VG, Cavalcoli J, Wright Z, Karnovsky A*, et al.* ConceptGen: a gene set enrichment and gene set relation mapping tool. *Bioinformatics* 2010, **26**(4)**:** 456-463.

85. Mok SC, Bonome T, Vathipadiekal V, Bell A, Johnson ME, Wong KK*, et al.* A gene signature predictive for outcome in advanced ovarian cancer identifies a survival factor: microfibril-associated glycoprotein 2. *Cancer cell* 2009, **16**(6)**:** 521-532.

86. Beroukhim R, Brunet JP, Di Napoli A, Mertz KD, Seeley A, Pires MM*, et al.* Patterns of gene expression and copy-number alterations in von-hippel lindau disease-associated and sporadic clear cell carcinoma of the kidney. *Cancer research* 2009, **69**(11)**:** 4674-4681.

87. Beleut M, Zimmermann P, Baudis M, Bruni N, Buhlmann P, Laule O*, et al.* Integrative genome-wide expression profiling identifies three distinct molecular subgroups of renal cell carcinoma with different patient outcome. *BMC cancer* 2012, **12:** 310.

88. Pena-Llopis S, Vega-Rubin-de-Celis S, Liao A, Leng N, Pavia-Jimenez A, Wang S*, et al.* BAP1 loss defines a new class of renal cell carcinoma. *Nature genetics* 2012, **44**(7)**:** 751-759.

89. Kikuchi A, Ishikawa T, Mogushi K, Ishiguro M, Iida S, Mizushima H*, et al.* Identification of NUCKS1 as a colorectal cancer prognostic marker through integrated expression and copy number analysis. *International journal of cancer Journal international du cancer* 2013, **132**(10)**:** 2295-2302.

90. Lu TP, Lai LC, Tsai MH, Chen PC, Hsu CP, Lee JM*, et al.* Integrated analyses of copy number variations and gene expression in lung adenocarcinoma. *PloS one* 2011, **6**(9)**:** e24829.

91. Hook KE, Garza SJ, Lira ME, Ching KA, Lee NV, Cao J*, et al.* An integrated genomic approach to identify predictive biomarkers of response to the aurora kinase inhibitor PF-03814735. *Molecular cancer therapeutics* 2012, **11**(3)**:** 710-719.

92. Chung CH, Ely K, McGavran L, Varella-Garcia M, Parker J, Parker N*, et al.* Increased epidermal growth factor receptor gene copy number is associated with poor prognosis in head and neck squamous cell carcinomas. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2006, **24**(25)**:** 4170-4176.

93. Day RS, McDade KK, Chandran UR, Lisovich A, Conrads TP, Hood BL*, et al.* Identifier mapping performance for integrating transcriptomics and proteomics experimental results. *BMC bioinformatics* 2011, **12:** 213.

94. Freeman TJ, Smith JJ, Chen X, Washington MK, Roland JT, Means AL*, et al.* Smad4-mediated signaling inhibits intestinal neoplasia by inhibiting expression of beta-catenin. *Gastroenterology* 2012, **142**(3)**:** 562-571 e562.

95. Kao KJ, Chang KM, Hsu HC, Huang AT. Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. *BMC cancer* 2011, **11:** 143.

96. Micke P, Edlund K, Holmberg L, Kultima HG, Mansouri L, Ekman S*, et al.* Gene copy number aberrations are associated with survival in histologic subgroups of non-small cell lung cancer. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* 2011, **6**(11)**:** 1833-1840.

97. Urquidi V, Goodison S, Cai Y, Sun Y, Rosser CJ. A candidate molecular biomarker panel for the detection of bladder cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2012, **21**(12)**:** 2149-2158.

98. Riester M, Taylor JM, Feifer A, Koppie T, Rosenberg JE, Downey RJ*, et al.* Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2012, **18**(5)**:** 1323-1333.

99. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R*, et al.* Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer research* 2012, **72**(1)**:** 100-111.

100. Botling J, Edlund K, Lohr M, Hellwig B, Holmberg L, Lambe M*, et al.* Biomarker discovery in

non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2013, **19**(1)**:** 194-204.

101. Jung AC, Job S, Ledrappier S, Macabre C, Abecassis J, De Reynies A*, et al.* A poor prognosis subtype of HNSCC is consistently observed across methylome, transcriptome and miRNome analysis. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2013.

102. Madhavan S, Zenklusen JC, Kotliarov Y, Sahni H, Fine HA, Buetow K. Rembrandt: helping personalized medicine become a reality through integrative translational research. *Molecular cancer research : MCR* 2009, **7**(2)**:** 157-167.