Supplementary Information to

INFORMATION CONTENT AND ANALYSIS METHODS FOR MULTI-MODAL HIGH-THROUGHPUT BIOMEDICAL DATA

Bisakha Ray, Mikael Henaff, Sisi Ma, Efstratios Efstathiadis, Eric R. Peskin,
Marco Picone, Tito Poli, Constantin F. Aliferis, Alexander Statnikov

---

Table S1: Data modalities, assaying platforms, and number of variables for each dataset.

TCGA_BRCA1 and **TCGA_BRCA2** Datasets		
Modality	Assaying platform information	Number of variables
Gene expression	AgilentG4502A_07 (TCGA code)	90,797
DNA methylation	HumanMethylation27 (TCGA code) or HumanMethylation450 (TCGA code)	27,578 or 485,575
Protein expression	MDA_RPPA_Core (TCGA code)	166
Somatic mutations	IlluminaGA_DNASeq (TCGA code)	12,497
Clinical	(see text and **Table S2** for details)	24
TCGA_OVCA Datasets		
Modality	Assaying platform information	Number of variables
Gene expression	AgilentG4502A_07 (TCGA code)	90,797
DNA methylation	HumanMethylation27 (TCGA code)	27,578
Protein expression	MDA_RPPA_Core (TCGA code)	166
miRNA expression	H-miRNA_8x15Kv2 (TCGA code)	2,423
Clinical	(see text and **Table S2** for details)	23
MSKCC_PRCA Datasets		
Modality	Assaying platform information	Number of variables
Gene expression	Affymetrix Human Exon 1.0 ST Array	26,447
Copy number	Agilent-014693 Human Genome CGH Microarray 244A	18,202
miRNA expression	Agilent-019118 Human miRNA Microarray 2.0 G4470B	368
Clinical	(see text and **Table S2** for details)	9
NEOMARK Datasets		
Modality	Assaying platform information	Number of variables
Gene expression	Agilent	25,702
Tumor imaging	MR/CT	34
Clinical	(see text and **Table S2** for details)	48
METABRIC Datasets		
Modality	Assaying platform information	Number of variables
Gene expression	Illumina HT 12	48,803
GWAS	Affymetrix SNP 6.0	909,662
Clinical	(see text and **Table S2** for details)	16

Table S2: Clinical predictors for each dataset. 'N' is numeric/continuous and 'B' is binary in the "Value" column.

TCGA_BRCA1 and **TCGA_BRCA2** Datasets			
Predictor	Value	Predictor	Value
Age at initial pathologic diagnosis	**N**	Her2 immunohistochemistry level result = 0+/1+	**B**
Anatomic organ subdivision = Left	**B**	Her2 immunohistochemistry level result = 2+	**B**
Anatomic organ subdivision = Right	**B**	Her2 immunohistochemistry level result = 3+	**B**
Anatomic organ subdivision = Not Specified	**B**	Menopause status = Not Specified or Indeterminate or Perimenopausal (grouped because of rare values)	**B**
Ethnicity = Non-Hispanic	**B**	Menopause status = Premenopausal	**B**
Ethnicity = Hispanic	**B**	Menopause status = Postmenopausal	**B**
Ethnicity = Not Specified	**B**	Prior diagnosis = Yes/No	**B**
Her2/Neu immunohistochemistry receptor status = Not Specified	**B**	Race = Not Specified	**B**
Her2/Neu immunohistochemistry receptor status = Negative	**B**	Race = White	**B**
Her2/Neu immunohistochemistry receptor status = Positive	**B**	Race = Black or African American	**B**
Her2/Neu immunohistochemistry receptor status = Indeterminate/Equivocal	**B**	Race = Asian	**B**
Her2 immunohistochemistry level result = Not Specified	**B**	Gender = Female/Male	**B**
TCGA_OVCA Datasets			
Predictor	Value	Predictor	Value
Age at initial pathologic diagnosis	**N**	Karnofsky score = 80	**B**
Anatomic organ subdivision = Left	**B**	Karnofsky score = 100	**B**
Anatomic organ subdivision = Right	**B**	Pretreatment history = Not Specified	**B**
Anatomic organ subdivision = Not Specified	**B**	Pretreatment history = Present	**B**
Ethnicity = Non-Hispanic	**B**	Pretreatment history = Absent	**B**
Ethnicity = Hispanic	**B**	Race = Not Specified	**B**
Ethnicity = Not Specified	**B**	Race = White	**B**
Jewish origin = Missing	**B**	Race = Black or African American	**B**
Jewish origin = Ashkenazi	**B**	Race = Asian	**B**
Jewish origin = Non-Ashkenazi	**B**	Race = American Indian	**B**
Karnofsky score = Not Specified	**B**	Gender = Female/Male	**B**
Karnofsky score = 60	**B**		
MSKCC_PRCA Datasets			
Predictor	Value	Predictor	Value
Age at diagnosis	**N**	Race = White Hispanic	**B**
Prostate specific antigen at diagnosis	**N**	Race = Asian	**B**
Race = Black Non-Hispanic	**B**	Race = Other	**B**
Race = White Non-Hispanic	**B**	Race = Not Specified	**B**
Race = Black Hispanic	**B**		

(Table S2 is continued on the next page)

(Table S2 is continued from the next page)

NEOMARK Datasets			
Predictor	Value	Predictor	Value
Weight	**N**	BMI	**N**
Weight = Not Available	**B**	BMI = Not Specified	**B**
Height	**N**	Duration of precancerous lesions (months)	**N**
Height = Not Available	**B**	Duration of precancerous lesions (months)= Not Specified	**B**
Diabetes	**B**	Drinking habit = Heavy Drinker	**B**
Allergies	**B**	Drinking habit = Moderate Drinker	**B**
High cholesterol	**B**	Drinking habit = Not Specified	**B**
Hypertension	**B**	Mobile prosthesis = Inferior	**B**
Familiar history of malignance	**B**	Mobile prosthesis = Inferior and superior	**B**
Smoker	**B**	Mobile prosthesis = Not Specified	**B**
Cigarette quantity per day	**N**	Mobile prosthesis = Superior	**B**
Cigarette quantity per day = Not Specified	**B**	Oral hygiene = Adequate	**B**
Years smoking	**N**	Oral hygiene = Not Adequate	**B**
Years smoking = Not Specified	**B**	Oral hygiene = Not Specified	**B**
Ex-smoker	**B**	Eating habit = Normal	**B**
When quit smoking (years ago)	**N**	Eating habit = Not Normal	**B**
When quit smoking (years ago) = Not Specified	**B**	Eating habit = Not Available	**B**
Alcohol use	**B**	Substance exposition = No	**B**
Mechanical trauma	**B**	Substance exposition = Not Specified	**B**
Dental cusps	**B**	Precancerous lesions = Erythroplankia	**B**
Galvanic current	**B**	Precancerous lesions = Leukoplankia	**B**
Infection	**B**	Precancerous lesions = No Lesions	**B**
Hb hematic concentration	**N**	Precancerous lesions = Not Specified	**B**
Hb hematic concentration = Not Specified	**B**	Precancerous lesions = Oral Submucosus Fibrosis	**B**
METABRIC Datasets			
Predictor	Value	Predictor	Value
Age at initial pathologic diagnosis	**N**	Site = 3	**B**
Menopausal status = pre	**B**	Treatment = CT	**B**
Menopausal status = post	**B**	Treatment = CT/HT	**B**
P53 mutation status = mutated	**B**	Treatment = CT/HT/RT	**B**
P53 mutation status = wild-type	**B**	Treatment = HT	**B**
P53 mutation status = Not Specified	**B**	Treatment = HT/RT	**B**
Site = 1	**B**	Treatment = RT	**B**
Site = 2	**B**	Treatment = NONE	**B**

Table S3: Core methods used in this work.

<table>
<tr><th colspan="2">STANDARD UNI-MODAL METHODS</th></tr>
<tr><th>Classification algorithm</th><th>Feature selection algorithm</th></tr>
<tr><td>• Support vector machines (SVMs) with linear kernel: default penalty parameter C [1]
• SVMs with polynomial kernel: penalty parameter C and kernel degree q selected by nested cross-validation [1]
• Random forests: default parameters [2]
• Bayesian logistic regression: Gaussian priors, variance parameter selected by cross-validation [3,4]
• Kernel ridge regression: radial basis function kernel, ridge and gamma parameters selected by cross-validation [5-7]</td><td>○ Use all available features from single data modality; no feature selection is performed
○ Perform feature selection in single data modality; use SVM-RFE [8]</td></tr>
<tr><th colspan="2">MULTI-MODAL UNIFORM (MMU) METHODS</th></tr>
<tr><td colspan="2">Each classifier is trained on features from multiple modalities obtained by different feature selection methods. Data modalities are treated uniformly.</td></tr>
<tr><th>Classification algorithm</th><th>Feature selection algorithm</th></tr>
<tr><td>• Support vector machines (SVMs) with linear kernel: default penalty parameter C [1]
• Support vector machines (SVMs) with polynomial kernel: penalty parameter C and kernel degree q selected by nested cross-validation [1]
• Random forests: default parameters [2]
• Bayesian logistic regression: Gaussian priors, variance parameter selected by cross-validation [3,4]
• Kernel ridge regression: radial basis function kernel, ridge and gamma parameters selected by cross-validation [5-7]</td><td>○ Use all available features from all data modalities; no feature selection is performed (the datasets corresponding to different modalities are simply “concatenated”)
○ Perform feature selection independently for each data modality, then return the union of selected features; use SVM-RFE [8] on the dataset from each data modality
○ Perform feature selection in all data modalities at once; i.e., use SVM-RFE [8] on the dataset obtained by “concatenating” all data modalities).</td></tr>
<tr><th colspan="2">MULTI-MODAL ENSEMBLE (MME) METHODS</th></tr>
<tr><td colspan="2">The following methods ensemble classification models derived from individual data modalities.
• Ensemble of random forests (average): A random forest is fit independently for each data modality, then an ensemble classifier is constructed by averaging the probabilities output by random forests
• Ensemble of random forests (max): A random forest is fit independently for each data modality, then an ensemble classifier is constructed by taking the maximum of the probabilities output by random forests
• Ensemble of SVMs w/o feature selection: A linear SVM model is fit independently for each data modality without feature selection, then an ensemble classifier is constructed by logistic regression using the predictions for each modality as input features
• Ensemble of SVMs with feature selection: A linear SVM model is fit independently for each data modality with feature selection by SVM-RFE, then an ensemble classifier is constructed by logistic regression using the predictions for each modality as input features</td></tr>
<tr><th colspan="2">MULTI-MODAL SPECIFIC (MMS) METHODS</th></tr>
<tr><td colspan="2">The following methods are classifiers designed specifically to handle multi-modal data.
• Sequential minimal optimization (SMO) for multiple kernel learning [9,10]: Linear kernel
• SMO for multiple kernel learning [9,10]: Linear normalized kernel
• SMO for multiple kernel learning [9,10]: Polynomial kernel
• SMO for multiple kernel learning [9,10]: Polynomial normalized kernel
• Least-squares SVMs for high-dimensional multi-modal data [11]: Normalized linear kernel, fixed weights, using all features</td></tr>
</table>

Table S4: Parameters and software implementations of core methods.

Method name	Parameter	Value(s)	Software implementation
GENERAL-PURPOSE CLASSIFICATION ALGORITHMS (USED BOTH IN STANDARD UNI-MODAL METHODS AND MULTI-MODAL UNIFORM (MMU) METHODS)			
SVMs with linear kernel	*C (error penalty)*	1	libsvm [12,13] http://www.csie.ntu.edu.tw/~cjlin/libsvm
SVMs with polynomial kernel	*C (error penalty)*	optimized over (0.01, 0.1, 1, 10, 100)	
	q (polynomial degree)	optimized over (1, 2, 3)	
Kernel ridge regression	*Ridge*	optimized over (10^{-10}, 10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3},10^{-2},10^{-1},1)	clop [7] http://clopinet.com/isabelle/Projects/ETH/Feature_Selection_w_CLOP.html
	γ (kernel width)	optimized over (0.01, 0.1, 1, 10, 100) / total number of features	
Bayesian logistic regression	*Prior*	Gaussian	Bbr http://www.bayesianregression.org/
	variance parameter	optimized over (2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 1, 2, 2^2, 2^3, 2^4)	
Random forests	*ntree (number of trees)*	500	R package randomForest http://cran.r-project.org/web/packages/randomForest/index.html
	mtry (number of features in a tree)	$\sqrt{\text{total number of features}}$	
	nodesize (min. size of terminal nodes of a tree)	1	
GENERAL-PURPOSE FEATURE SELECTION ALGORITHMS (USED BOTH IN STANDARD UNI-MODAL METHODS AND MULTI-MODAL UNIFORM (MMU) METHODS)			
No feature selection	-	-	-
SVM-RFE	*proportion of features to discard at each iteration in order to create nested feature subsets*	20%	internal implementation on top of libsvm
MULTI-MODAL ENSEMBLE (MME) METHODS			
Ensemble of random forests (average)	*Random forest parameters*	see above for Random forests	internal implementation on top of R package randomForest
	Ensemble function	average	
Ensemble of random forests (max)	*Random forest parameters*	see above for Random forests	
	Ensemble function	max	
Ensemble of SVMs w/o feature selection	*SVM parameters*	see above for SVMs with linear kernel	internal implementation on top of libsvm and R function glm
	Ensemble function	obtained by logistic regression	
Ensemble of SVMs with feature selection	*SVM parameters*	see above for SVMs with linear kernel	
	Feature selection parameters	see above for SVM-RFE	
MULTI-MODAL SPECIFIC (MMS) METHODS			
SMO for multiple kernel learning	*kernel*	linear	SKMsmo http://www.di.ens.fr/~obozinski/code.html
SMO for multiple kernel learning	*kernel*	linear normalized	
SMO for multiple kernel learning	*kernel*	polynomial, degree = 2	
SMO for multiple kernel learning	*kernel*	polynomial, degree = 2, normalized	
Least-squares SVMs for high-dimensional multi-modal data	*kernel*	linear normalized	HDIDIT http://homes.esat.kuleuven.be/~bioiuser/HIDIDIT/
	weights	fixed	
	feature selection	none	

Table S5: Prior use of SVM-RFE feature selection method in various data modalities.

Modality	Study references
Clinical	[14,15]
Gene Expression	[16-19]
Protein Expression and Proteomics	[20-23]
Somatic Mutations	[24]
DNA Methylation	[25,26]
miRNA Expression	[27-29]
Copy Number	[30,31]
Tumor Imaging	[32-35]
GWAS	[36-38]

Table S6: Comparison of 16 feature selection methods with SVM-RFE in TCGA_BRCA1 datasets/predictive tasks. The first two tables report predictive performance (AUC) averaged over 8 TCGA_BRCA1 datasets/predictive tasks for various feature selection methods. As can be seen, SVM-RFE yields higher average predictive performance than any of 16 other feature selection methods, both in uni-modal and multi-modal applications.

Mean predictive performance (AUC) obtained by linear SVM for various feature selection methods

	MRMR-50	MRMR-100	MRMR-500	MRMR-1000	SPCA-50	SPCA-100	SPCA-500	SPCA-1000	UKW-50	UKW-100	UKW-500	UKW-1000	UKW-Alpha	UKW-FDR	UKW	MRMR	SVM-RFE
Uni-Modal Clinical	0.501	0.501	0.501	0.501	0.501	0.501	0.501	0.501	0.501	0.501	0.501	0.501	0.516	0.515	0.493	0.528	*0.640*
Uni-Modal Gene Expression	0.617	0.613	0.614	0.624	0.626	0.631	0.609	0.610	0.610	0.606	0.580	0.588	0.591	0.609	0.640	0.644	*0.758*
Uni-Modal Protein Expression	0.680	0.687	0.689	0.689	0.619	0.638	0.689	0.689	0.668	0.668	0.689	0.689	0.628	0.644	0.623	0.638	*0.789*
Uni-Modal Somatic Mutations	0.555	0.581	0.574	0.572	0.507	0.499	0.523	0.524	0.567	0.562	0.571	0.557	0.580	0.583	0.620	0.628	*0.700*
Uni-Modal DNA Methylation	0.622	0.615	0.609	0.605	0.544	0.554	0.573	0.593	0.588	0.585	0.585	0.586	0.585	0.614	0.616	0.619	*0.750*
Multi-Modal Uniform*	0.617	0.612	0.609	0.617	0.592	0.599	0.609	0.607	0.614	0.607	0.587	0.598	0.598	0.613	0.628	0.629	*0.753*
Multi-Modal Uniform†	0.624	0.636	0.618	0.623	0.628	0.618	0.612	0.608	0.600	0.600	0.604	0.604	0.598	0.651	0.674	0.664	*0.762*

Mean predictive performance (AUC) obtained by Bayesian Logistic Regression for various feature selection methods

	MRMR-50	MRMR-100	MRMR-500	MRMR-1000	SPCA-50	SPCA-100	SPCA-500	SPCA-1000	UKW-50	UKW-100	UKW-500	UKW-1000	UKW-Alpha	UKW-FDR	UKW	MRMR	SVM-RFE
Uni-Modal Clinical	0.519	0.519	0.519	0.519	0.518	0.518	0.505	0.518	0.505	0.518	0.518	0.518	0.546	0.526	0.511	0.540	*0.555*
Uni-Modal Gene Expression	0.614	0.611	0.619	0.637	0.629	0.624	0.554	0.632	0.611	0.602	0.591	0.595	0.599	0.610	0.641	0.637	*0.728*
Uni-Modal Protein Expression	0.665	0.671	0.669	0.669	0.604	0.620	0.626	0.669	0.659	0.657	0.668	0.694	0.633	0.644	0.635	0.637	*0.744*
Uni-Modal Somatic Mutations	0.550	0.577	0.561	0.573	0.501	0.491	0.514	0.525	0.573	0.566	0.570	0.571	0.599	0.583	0.613	0.617	*0.683*
Uni-Modal DNA Methylation	0.613	0.619	0.608	0.607	0.542	0.544	0.569	0.583	0.592	0.599	0.605	0.605	0.624	0.613	0.634	0.608	*0.719*
Multi-Modal Uniform*	0.616	0.616	0.618	0.621	0.633	0.601	0.614	0.613	0.611	0.606	0.594	0.609	0.552	0.613	0.633	0.619	*0.731*
Multi-Modal Uniform†	0.630	0.637	0.626	0.625	0.614	0.615	0.609	0.604	0.612	0.609	0.616	0.615	0.598	0.646	0.668	0.659	*0.759*

Description of used feature selection methods

MRMR-N (N=50,100,500,1000)	Minimum Redundancy Maximum Relevance Feature Selection (MRMR)[39] with up to N features selected#.
SPCA-N (N=50,100,500,1000)	Sparse Principal Component Analysis (SPCA)[40] with up to N features selected#.
UKW-N (N=50,100,500,1000)	Univariate Kruskal-Wallis (UKW)[41] with up to N features selected#.
UKW-Alpha	UKW with selecting features at 5% alpha level.
UKW-FDR	UKW with selecting features at 5% FDR level.
UKW	UKW with selecting the smallest subset of features that maximizes predictive performance of the classifier.
MRMR	MRMR with selecting the smallest subset of features that maximizes predictive performance of the classifier.
SVM-RFE	SVM-RFE (see **Table S4** for details).

* Multi-Modal Uniform approach was applied here with feature selection performed in all modalities at once.

† Multi-Modal Uniform approach was applied here with feature selection performed independently on individual modalities.

If N was larger than the number of features in a particular modality, all features from than modality were selected.

Table S7: Additional comparison of uni-modal and multi-modal approaches on datasets/predictive tasks that have gene expression and protein expression modalities (TCGA_BRCA1, TCGA_BRCA2, and TCGA_OVCA). The predictive performance (AUC) is averaged over 27 datasets/predictive tasks and is provided separately for linear SVM and Bayesian Logistic Regression (BLR) classifiers. As can be seen, using only gene expression and protein expression for multi-modal analyses does not improve average predictive performance compared to uni-modal approaches and multi-modal based on all 5 data modalities.

Approach	Details	Area under ROC curve (AUC)	
		Linear SVM	BLR
Uni-Modal Gene Expression	w/o feature selection	0.599	0.601
	with SVM-RFE feature selection	0.742	0.714
Uni-Modal Protein Expression	w/o feature selection	0.604	0.598
	with SVM-RFE feature selection	0.723	0.668
Multi-Modal Uniform with all 5 data modalities	w/o feature selection	0.602	0.580
	with features selected by SVM-RFE on all data modalities at once	0.735	0.717
	with features selected by SVM-RFE independently on each data modality	0.745	0.738
Multi-Modal Uniform with only Gene Expression and Protein Expression modalities	***w/o feature selection***	***0.600***	***0.595***
	with features selected by SVM-RFE on 2 data modalities at once	***0.710***	***0.683***
	with features selected by SVM-RFE independently on each data modality	***0.735***	***0.725***

Figure S1: Comparison #1: Uni-modal gene expression (GE) versus multi-modal uniform (MMU).

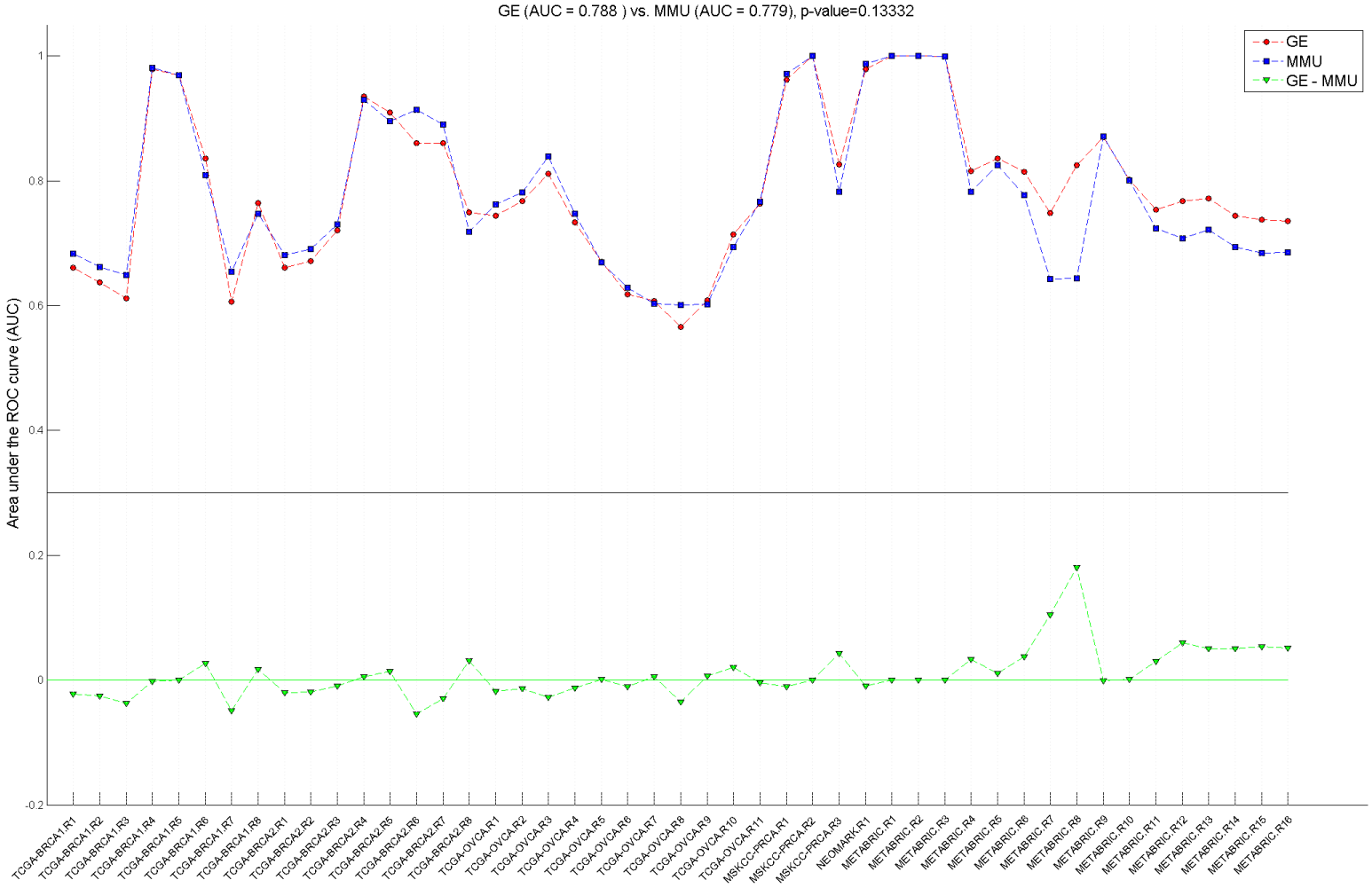


Figure S2: Comparison #2: Uni-modal gene expression (GE) versus multi-modal ensemble (MME).

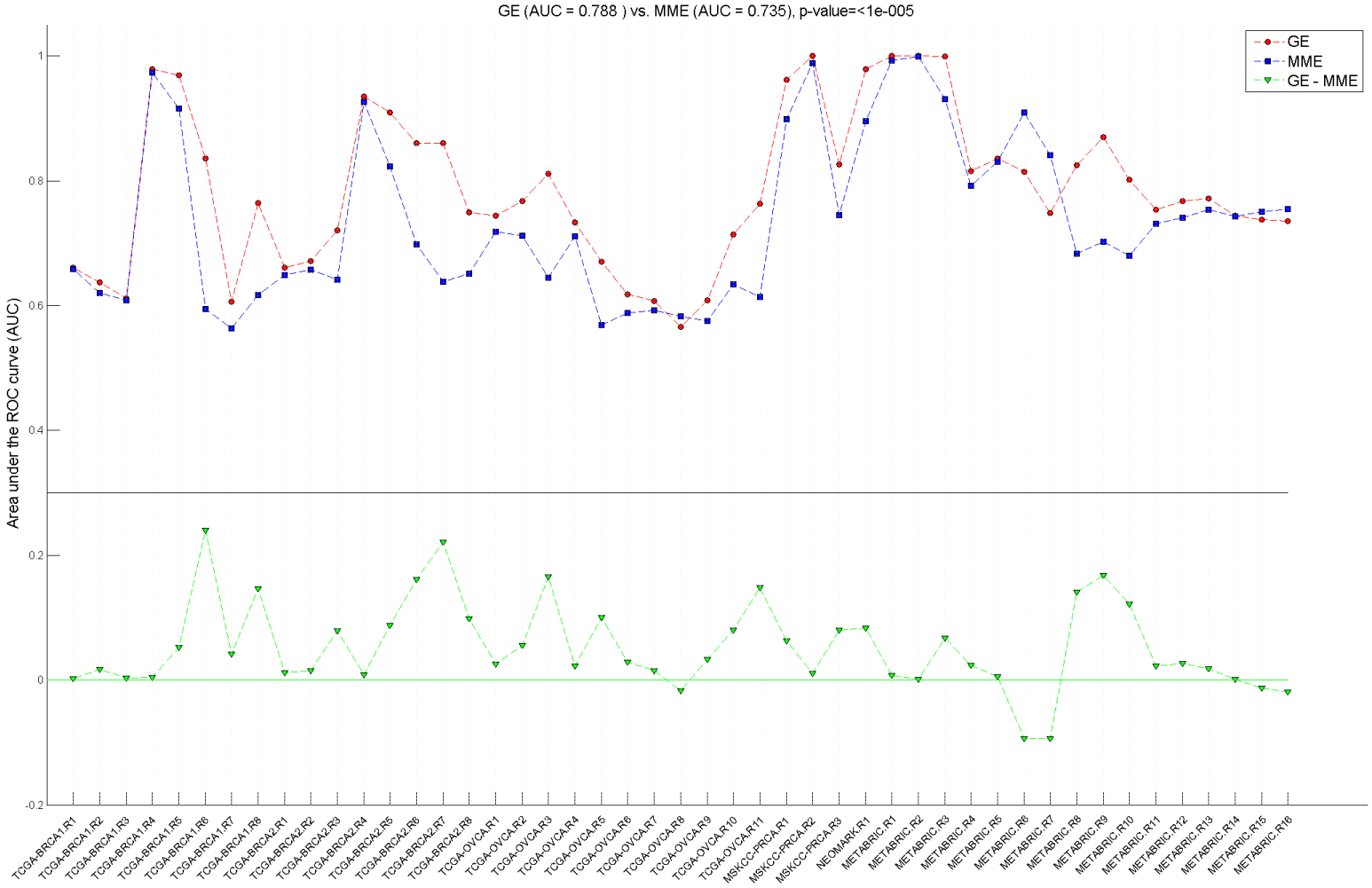


Figure S3: Comparison #3: Uni-modal gene expression (GE) versus multi-modal specific (MMS).

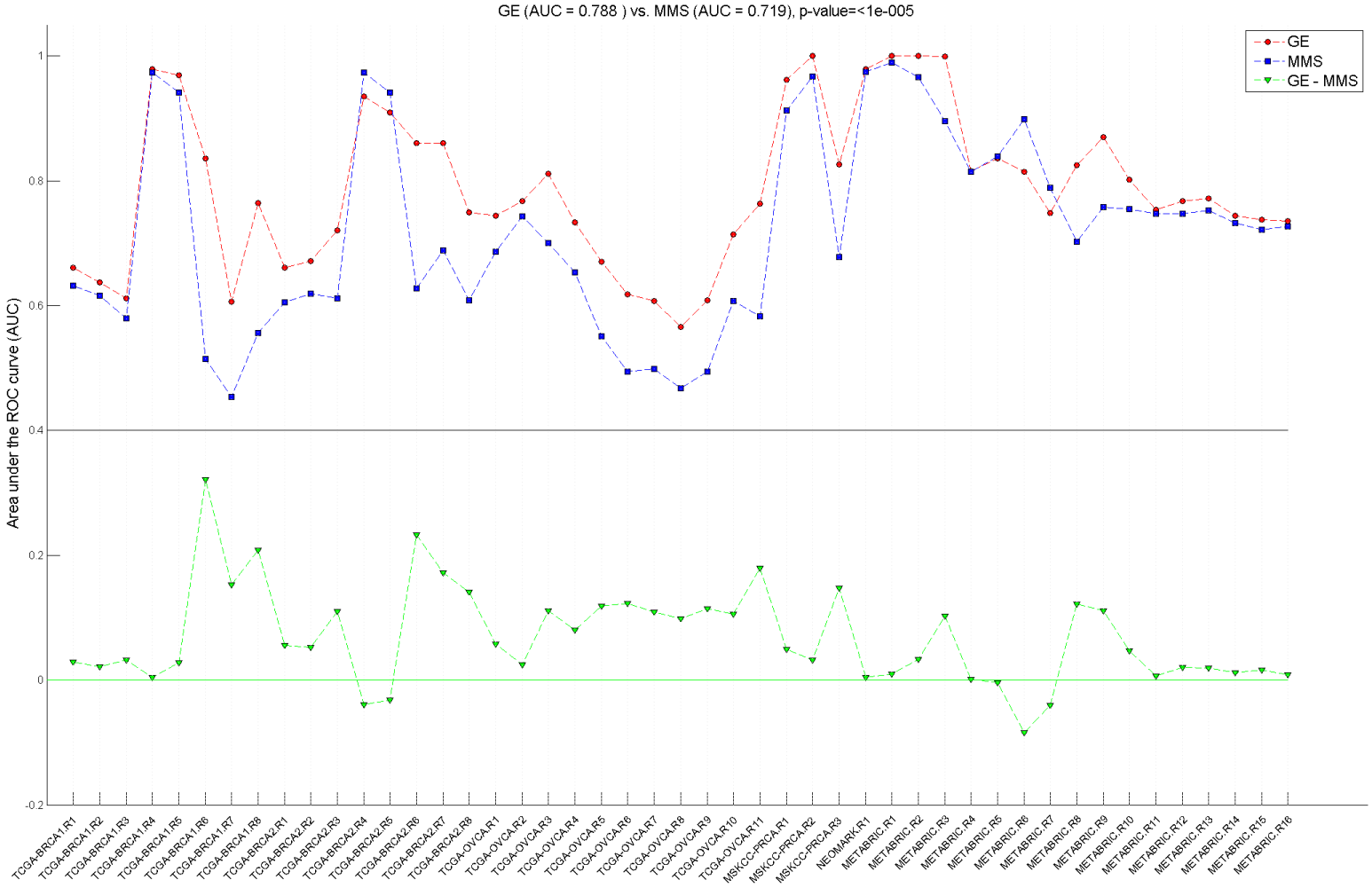


Figure S4: Comparison #4: Multi-modal uniform (MMU) versus multi-modal ensemble (MME).

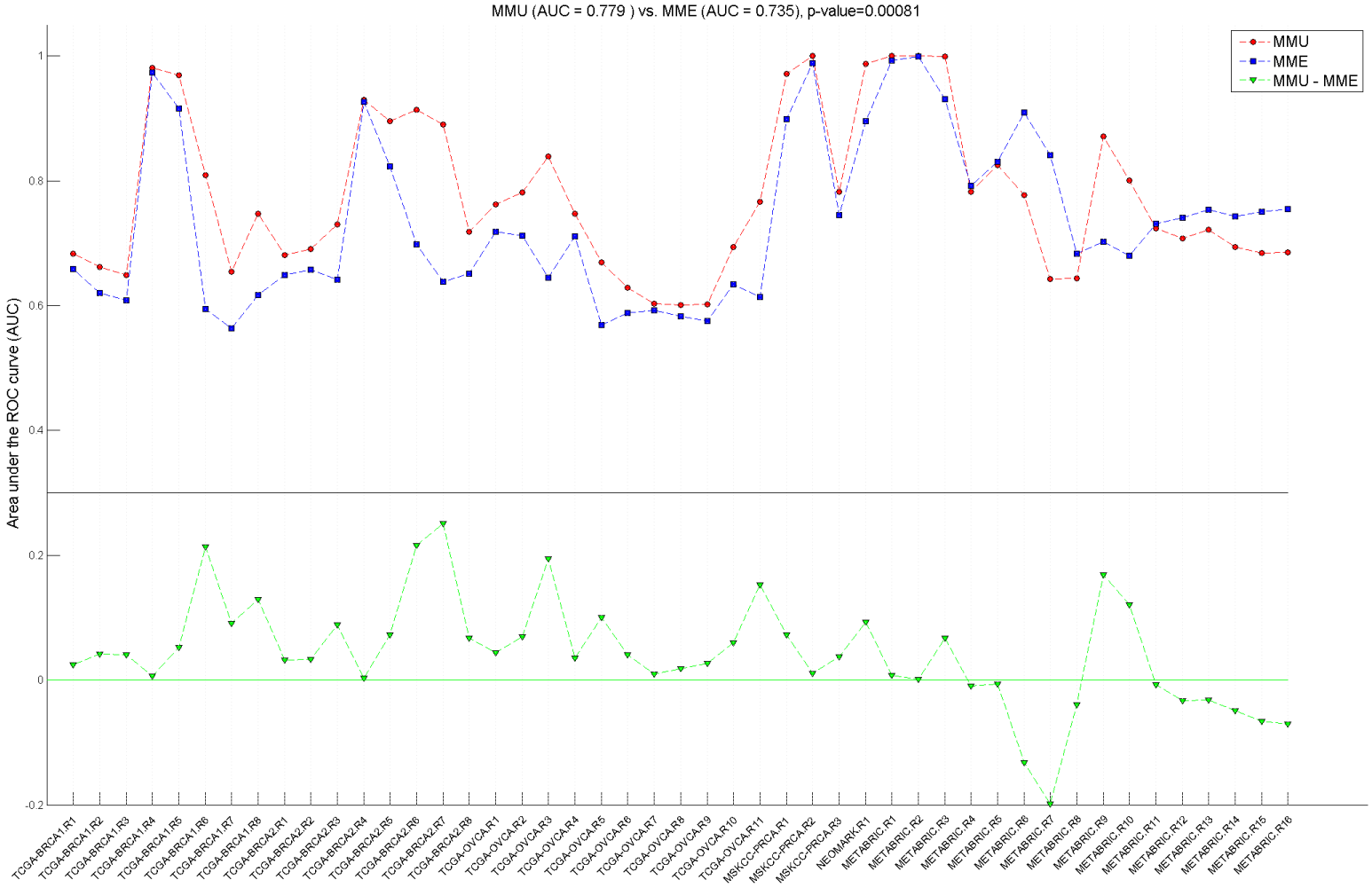


Figure S5: Comparison #5: Multi-modal uniform (MMU) versus multi-modal specific (MMS).

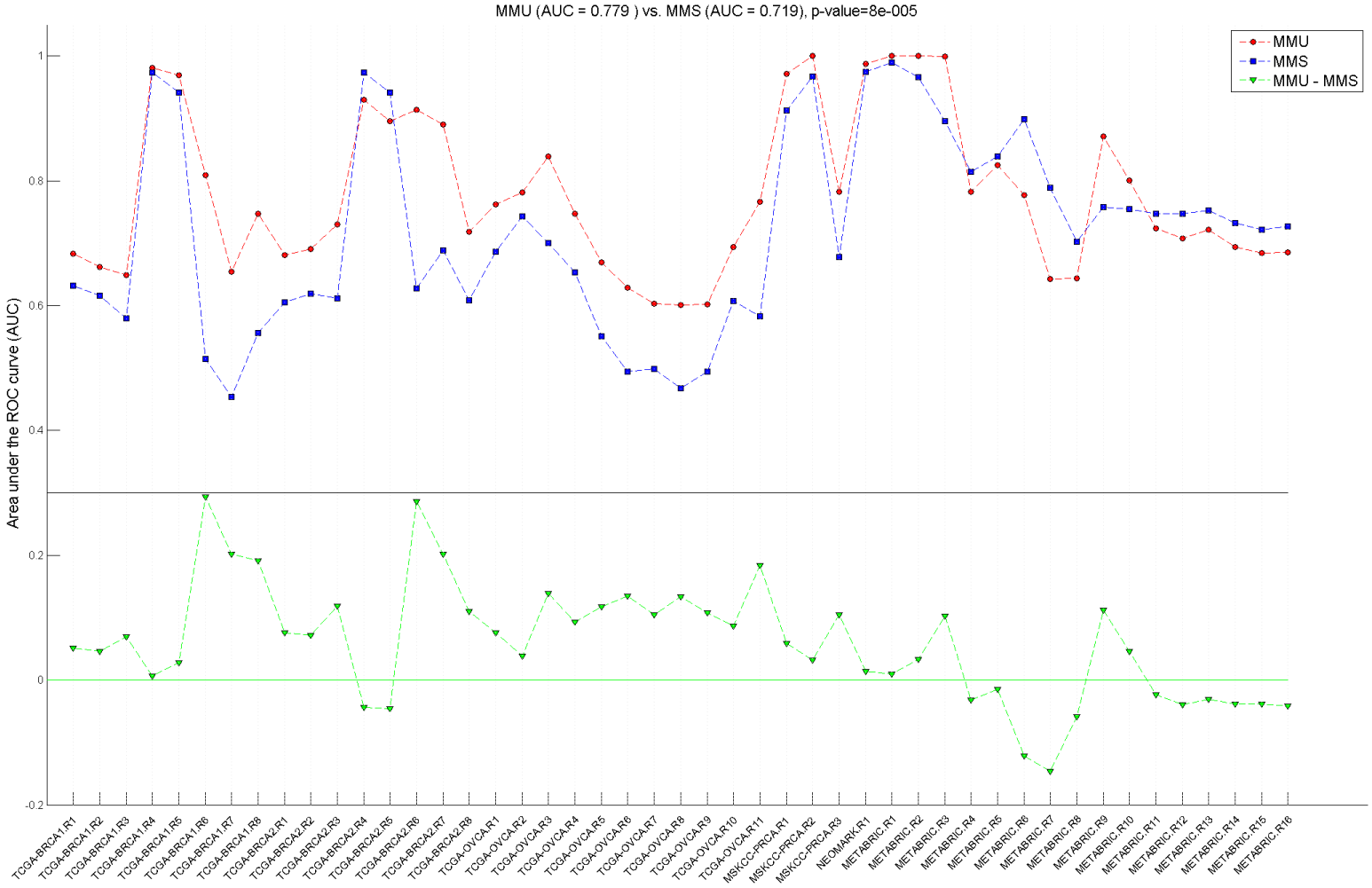
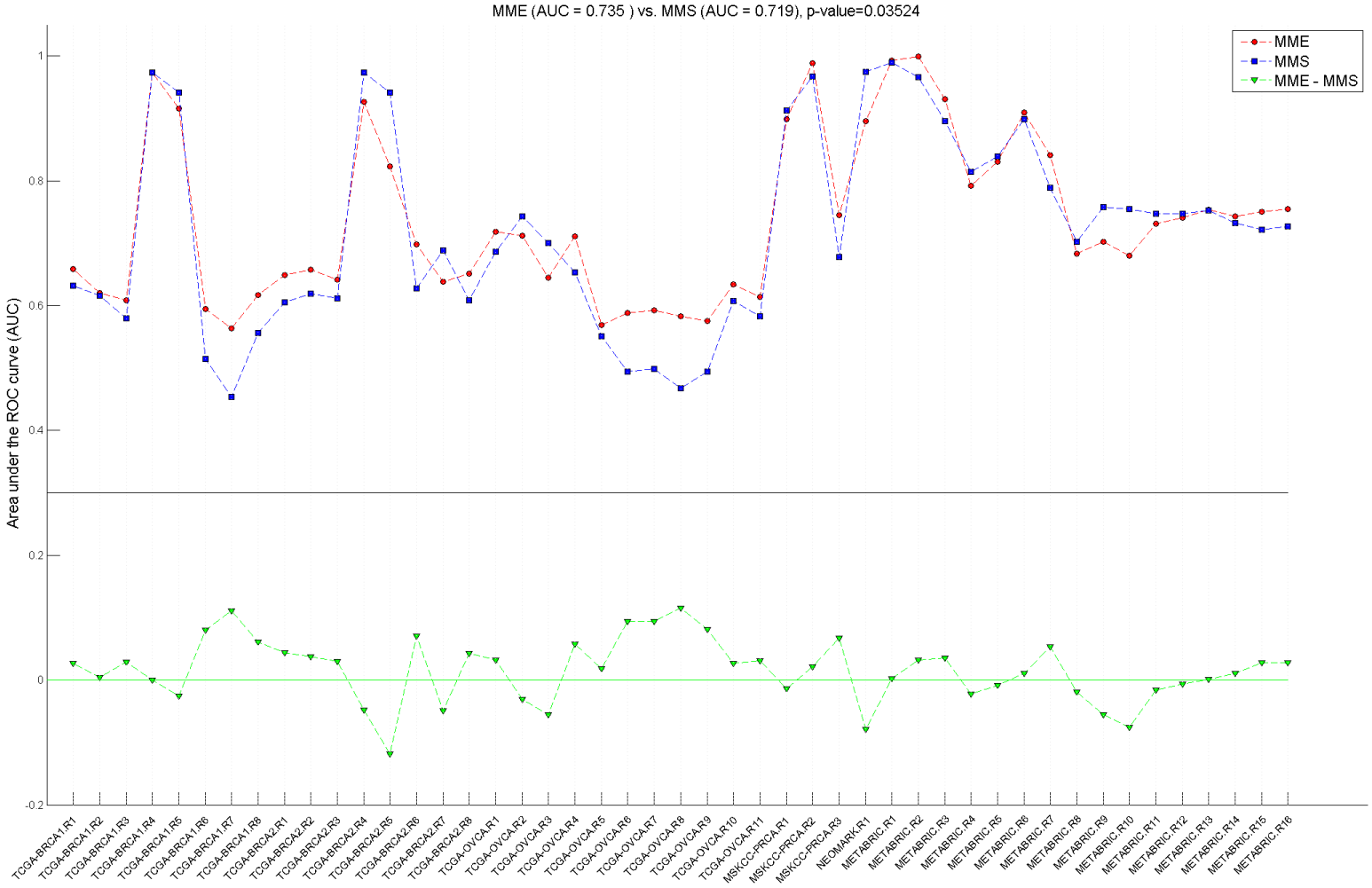


Figure S6: Comparison #6: Multi-modal ensemble (MME) versus multi-modal specific (MMS).



REFERENCES

1 Vapnik, V. N. *Statistical learning theory*. (Wiley, 1998).

2 Breiman, L. Random forests. *Machine Learning* **45**, 5-32 (2001).

3 Genkin, A., Lewis, D. D. & Madigan, D. Large-scale Bayesian logistic regression for text categorization. *Technometrics* **49**, 291-304 (2007).

4 Genkin, A., Lewis, D. D. & Madigan, D. Large-scale Bayesian logistic regression for text categorization. *Technical Report DIMACS* (2004).

5 Hastie, T., Tibshirani, R. & Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction*. (Springer, 2001).

6 Guyon, I. Kernel ridge regression tutorial. (http://clopinet.com/isabelle/Projects/ETH/KernelRidge.pdf, 2005).

7 Guyon, I. *et al.* Feature selection with the CLOP package. (Technical report, 2006, http://clopinet.com/isabelle/Projects/ETH/TM-fextract-class.pdf, 2006).

8 Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene selection for cancer classification using support vector machines. *Machine Learn* **46**, 389 - 422 (2002).

9 Lanckriet, G. R., Deng, M., Cristianini, N., Jordan, M. I. & Noble, W. S. Kernel-based data fusion and its application to protein function prediction in yeast. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 300-311 (2004).

10 Bach, F. R., Lanckriet, G. R. G. & Jordan, M. I. Fast kernel learning using sequential minimal optimization. (2004).

11 Daemen, A. *et al.* A kernel-based integration of genome-wide data for clinical decision support. *Genome medicine* **1**, 39, doi:10.1186/gm39 (2009).

12 Chang, C. C. & Lin, C. J. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* **2**, 27 (2011).

13 Fan, R. E., Chen, P. H. & Lin, C. J. Working set selection using second order information for training support vector machines. *Journal of Machine Learning Research* **6**, 1918 (2005).

14 Cho, B. H. *et al.* Application of irregular and unbalanced data to predict diabetic nephropathy using visualization and feature selection methods. *Artif Intell Med* **42**, 37-53, doi:DOI 10.1016/j.artmed.2007.09.005 (2008).

15 Yoon, S. & Kim, S. AdaBoost-based multiple SVM-RFE for classification of mammograms in DDSM. *Bmc Med Inform Decis* **9** (2009).

16 Mao, Y., Zhou, X. B., Pi, D. Y., Sun, Y. X. & Wong, S. T. C. Multiclass cancer classification by using fuzzy support vector machine and binary decision tree with gene selection. *J Biomed Biotechnol*, 160-171, doi:Doi 10.1155/Jbb.2005.160 (2005).

17 Li, Y., Campbell, C. & Tipping, M. Bayesian automatic relevance determination algorithms for classifying gene expression data. *Bioinformatics* **18**, 1332-1339, doi:DOI 10.1093/bioinformatics/18.10.1332 (2002).

18 Tang, Y. C., Zhang, Y. Q. & Huang, Z. FCM-SVM-RFE gene feature selection algorithm for leukemia classification from microarray gene expression data. *Ieee Int Conf Fuzzy*, 97-101 (2005).

19 Peng, S. H. *et al.* Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines. *Febs Lett* **555**, 358-362, doi:Doi 10.1016/S0014-5793(03)01275-4 (2003).

20 Carvalho, P. C., Hewel, J., Barbosa, V. C. & Yates, J. R. Identifying differences in protein expression levels by spectral counting and feature selection. *Genet Mol Res* **7**, 342-356 (2008).

21 Jong, K., Marchiori, E., Sebag, M. & van der Vaart, A. Feature selection in proteomic pattern data with support vector machines. *Proceedings of the 2004 Ieee Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 41-48 (2004).

22 Zhang, X. G. *et al.* Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data. *Bmc Bioinformatics* **7** (2006).

23 Rajapakse, J. C., Duan, K. B. & Yeo, W. K. Proteomic cancer classification with mass spectrometry data. *Am J Pharmacogenomic* **5**, 281-292, doi:Doi 10.2165/00129785-200505050-00001 (2005).

24 Kim, J., Gao, L. & Tan, K. Multi-Analyte Network Markers for Tumor Prognosis. *Plos One* **7** (2012).

25 Das, R. *et al.* Computational prediction of methylation status in human genomic sequences. *P Natl Acad Sci USA* **103**, 10713-10716, doi:DOI 10.1073/pnas.0602949103 (2006).
26 Meng, H. L., Murrelle, E. L. & Li, G. Y. Identification of a small optimal subset of CpG sites as bio-markers from high-throughput DNA methylation profiles. *Bmc Bioinformatics* **9** (2008).
27 Hu, X. X. *et al.* A MicroRNA Expression Signature for Cervical Cancer Prognosis. *Cancer Res* **70**, 1441-1448, doi:Doi 10.1158/0008-5472.Can-09-3289 (2010).
28 Peng, S. H., Zeng, X. M., Li, X. B., Peng, X. N. & Chen, L. B. Multi-class cancer classification through gene expression profiles: microRNA versus mRNA. *J Genet Genomics* **36**, 409-416, doi:Doi 10.1016/S1673-8527(08)60130-7 (2009).
29 Ahsen, M. E., Singh, N. K., Boren, T., Vidyasagar, M. & White, M. A. A New Feature Selection Algorithm for Two-Class Classification Problems and Application to Endometrial Cancer. *Ieee Decis Contr P*, 2976-2982 (2012).
30 Liu, J., Ranka, S. & Kahveci, T. Classification and feature selection algorithms for multi-class CGH data. *Bioinformatics* **24**, I86-I95, doi:DOI 10.1093/bioinformatics/btn145 (2008).
31 Zhang, J. X. *et al.* Analysis of bypass signaling in EGFR pathway and profiling of bypass genes for predicting response to anticancer EGFR tyrosine kinase inhibitors. *Mol Biosyst* **8**, 2645-2656, doi:Doi 10.1039/C2mb25165e (2012).
32 Zacharaki, E. I. *et al.* Classification of Brain Tumor Type and Grade Using MRI Texture and Shape in a Machine Learning Scheme. *Magn Reson Med* **62**, 1609-1618, doi:Doi 10.1002/Mrm.22147 (2009).
33 Tuia, D., Camps-Valls, G., Matasci, G. & Kanevski, M. Learning Relevant Image Features With Multiple-Kernel Classification. *Ieee T Geosci Remote* **48**, 3780-3791, doi:Doi 10.1109/Tgrs.2010.2049496 (2010).
34 De Martino, F. *et al.* Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage* **43**, 44-58, doi:DOI 10.1016/j.neuroimage.2008.06.037 (2008).
35 Archibald, R. & Fann, G. Feature selection and classification of hyperspectral images, with support vector machines. *Ieee Geosci Remote S* **4**, 674-677, doi:Doi 10.1109/Lgrs.2007.905116 (2007).
36 Yoon, D., Kim, Y. J. & Park, T. Phenotype prediction from genome-wide association studies: application to smoking behaviors. *Bmc Syst Biol* **6** (2012).
37 Kim, J., Sohn, I., Kim, D. & Jung, S. H. SNP Selection in Genome-Wide Association Studies via Penalized Support Vector Machine with MAX Test. *Comput Math Method M* (2013).
38 Augustin, R. *et al.* Bioinformatics identification of modules of transcription factor binding sites in Alzheimer's disease-related genes by in silico promoter analysis and microarrays. *International journal of Alzheimer's disease* **2011**, 154325, doi:10.4061/2011/154325 (2011).
39 Ding, C. & Peng, H. Minimum redundancy feature selection from microarray gene expression data. *J Bioinform.Comput.Biol.* **3**, 185-205 (2005).
40 Zou, H., Hastie, T. & Tibshirani, R. Sparse principal component analysis. *Journal of Computational and Graphical Statistics* **15**, 265-286 (2006).
41 Hollander, M. & Wolfe, D. *Nonparametric statistical methods*. Vol. 2nd (Wiley, 1999).