

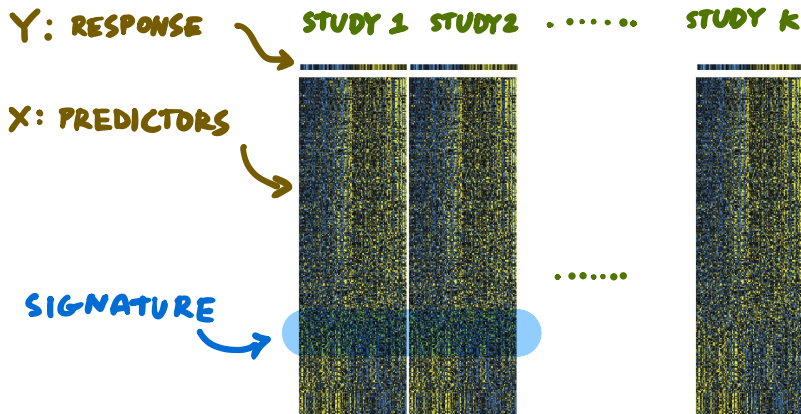
Validation and Replicability of Prediction Algorithms in Oncology

`giovanni_parmigiani@dfci.harvard.edu`

NATIONAL CANCER POLICY FORUM WORKSHOP 2018

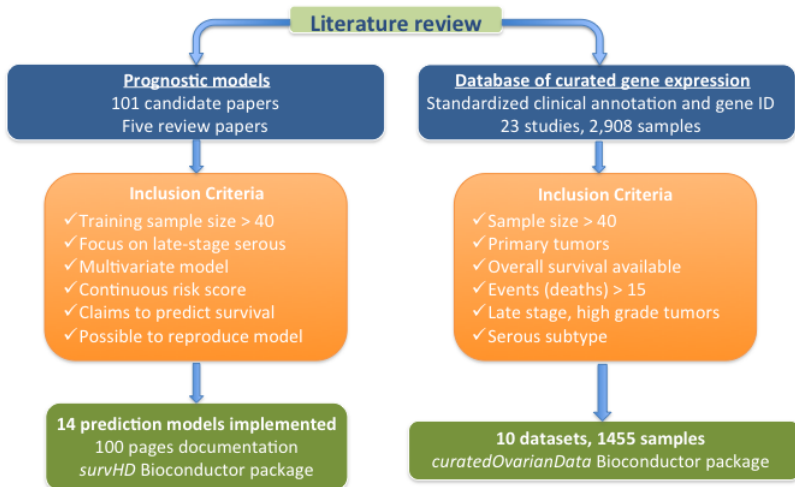
- Consultant: None Related to the Topic
- Speaker's Bureau: None
- Grant/Research support from: NIH-NCI; MIT-Harvard Bridge Project;
Licensing of BayesMendel software for genetic counseling.
- Stockholder in: CRA Health
- Honoraria from: Academic Only
- co-Founder / Board Member: Phaeno
- Relevant patents: POSTN for debulking in OC
- Employee of: Dana Farber Cancer Institute

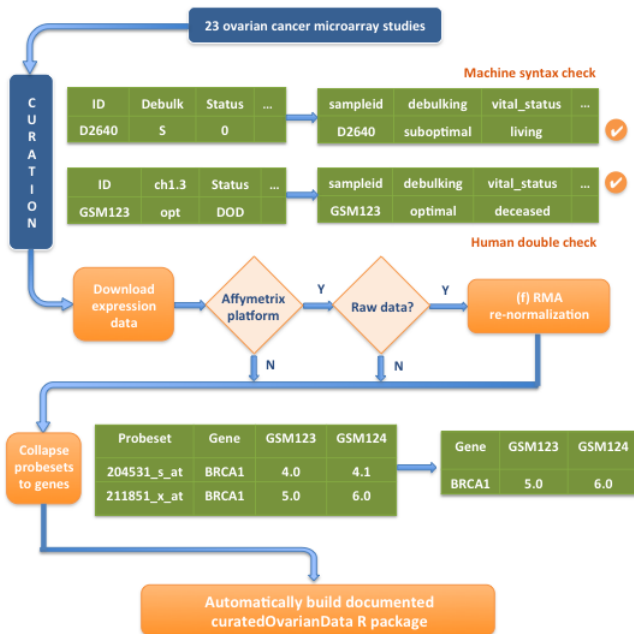
signatures, classifier, validation, replicability



CLASSIFIER: $\hat{Y} = f(\text{SIGNATURE})$

Meta-analysis overview





Model	Reproducibility†			
	Model Provided	Training Data Available	Validation Data Available	Verified Implementation
TCGA11 [12]	YES	YES	YES	YES
Denkert09 [13]	YES	YES	YES	YES
Bonome08_263genes [14]	YES	YES	YES	YES
Bonome08_572genes [14]	YES	YES	YES	YES
Mok09 [15]	NO	YES	YES	PARTIALLY
Yoshihara12 [16]	YES	-	YES	YES
Yoshihara10 [17]	YES	-	YES	YES
Bentink12 [18]	YES	-	YES	YES
Kang12 [19]	YES	YES	YES	PARTIALLY
Crijns09 [20]	NO	YES	NO	NO
Kernagis12 [21]	PARTIALLY	YES	YES	PARTIALLY
Sabatier11 [22]	PARTIALLY	NO	NO	NO
Konstantinopoulos10 [23]	YES	-	YES	PARTIALLY
Hernandez10 [24]	PARTIALLY	-	YES	PARTIALLY

Compare to *Ioannidis 08*: "We reproduced two analyses in principle and six partially or with some discrepancies; ten could not be reproduced."

Implemented Models

Validation Statistics for 14 Models in 10 Datasets

<i>Dataset Average</i>	1.81	1.47	1.43	1.41	1.39	1.37	1.35	1.14	1.11	1.04	
TCGA11	2.05	2.28	1.58	1.85	1.36	1.64	1.97	1.94	1.07	1.53	TCGA11
Yoshihara12	2.44	9.65	1.21	1.8	1.02	1.77	1.97	1.21	1.35	1.42	Yoshihara1
Yoshihara10	2.69	1.38	1.15	1.93	1.45	1.57	1.47	1.33	0.7	7.27	Yoshihara1
Kernagis12	2.65	1.39	2.91	2.08	1.45	1.39	1.25	1.23	1.32	0.87	Kernagis12
Crijns09	1.22	1.92	1.49	1.51	1.2	1.44	1.28	1.1	3.04	1.21	Crijns09
Bentink12	1.94	1.01	1.89	1.44	1.2	1.14	1.62	1.16	1.26	1.45	Bentink12
Bonome08_263genes	1.3	2.73	2	1.32	0.53	2.01	1.45	1.17	1.03	0.77	Bonome08
Mok09	1.54	1.82	3.18	1.71	0.89	1.58	1.28	0.98	0.95	1.39	Mok09
Bonome08_572genes	0.8	1.89	1.1	1.41	2.29	2.27	1.47	1.07	1.35	0.84	Bonome08
Sabatier11	1.95	1.15	1.17	1.41	1.72	1.07	1.19	1.11	1.3	0.73	Sabatier11
Denkert09	2.6	0.76	1.33	1.31	2.25	1.04	1.29	1.08	1.15	0.79	Denkert09
Kang12	2.14	1.19	0.81	0.85	1.21	1.46	1.17	1.55	1.02	0.73	Kang12
Konstantinopoulos10	1.34	1.01	0.82	1.07	2.05	1.07	1	1.15	0.97	1.09	Konstantinopoulos10
Hernandez10	0.68	0.55	1.07	0.71	0.86	1.21	0.79	1.04	0.9	1.03	Hernandez10

Expression Datasets

Dressman

Yoshihara 2012A

Mok

Tothill

Konstantinopoulos

Bonome

Bentink

TCGA

Crijns

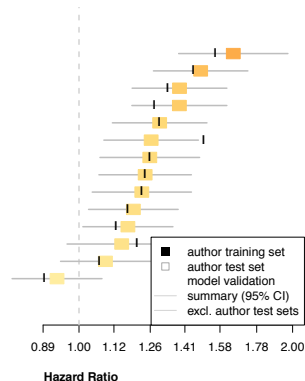
Yoshihara 2010

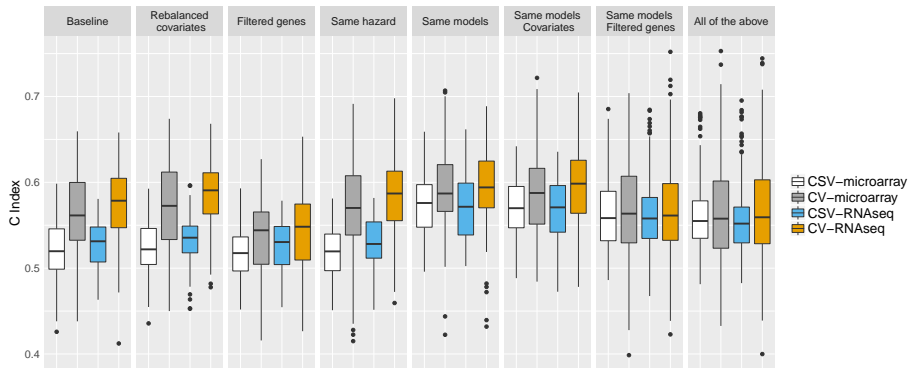
synthesis of validation matrix

Implemented Models

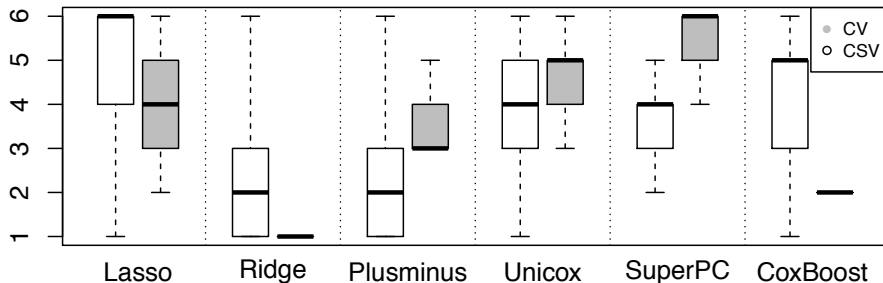
Validation Statistics for 14 Models in 10 Datasets

Dataset Average	1.81	1.47	1.43	1.41	1.39	1.37	1.35	1.14	1.11	1.04
TCGA11	2.05	2.28	1.58	1.85	1.36	1.64	1.97	1.94	1.07	1.53
Yoshihara12	2.44	9.65	1.21	1.8	1.02	1.77	1.97	1.21	1.35	1.42
Yoshihara10	2.69	1.38	1.15	1.93	1.45	1.57	1.47	1.33	0.7	7.27
Kernagis12	2.65	1.39	2.91	2.08	1.45	1.39	1.25	1.23	1.32	0.87
Crijns09	1.22	1.92	1.49	1.51	1.2	1.44	1.28	1.1	3.04	1.21
Bentink12	1.94	1.01	1.89	1.44	1.2	1.14	1.62	1.16	1.26	1.45
Bonome08_263genes	1.3	2.73	2	1.32	0.53	2.01	1.45	1.17	1.03	0.77
Mok09	1.54	1.82	3.18	1.71	0.89	1.58	1.28	0.98	0.95	1.39
Bonome08_572genes	0.8	1.89	1.1	1.41	2.29	2.27	1.47	1.07	1.35	0.84
Sabatier11	1.95	1.15	1.17	1.41	1.72	1.07	1.19	1.11	1.3	0.73
Denkert09	2.6	0.76	1.33	1.31	2.25	1.04	1.29	1.08	1.15	0.79
Kang12	2.14	1.19	0.81	0.85	1.21	1.46	1.17	1.55	1.02	0.73
Konstantinopoulos10	1.34	1.01	0.82	1.07	2.05	1.07	1	1.15	0.97	1.09
Hernandez10	0.68	0.55	1.07	0.71	0.86	1.21	0.79	1.04	0.9	1.03
Expression Datasets										
Dressman										
Yoshihara 2012A										
Mok										
Totthill										
Konstantinopoulos										
Bonome										
Bentink										
TCGA										
Crijns										
Yoshihara 2010										

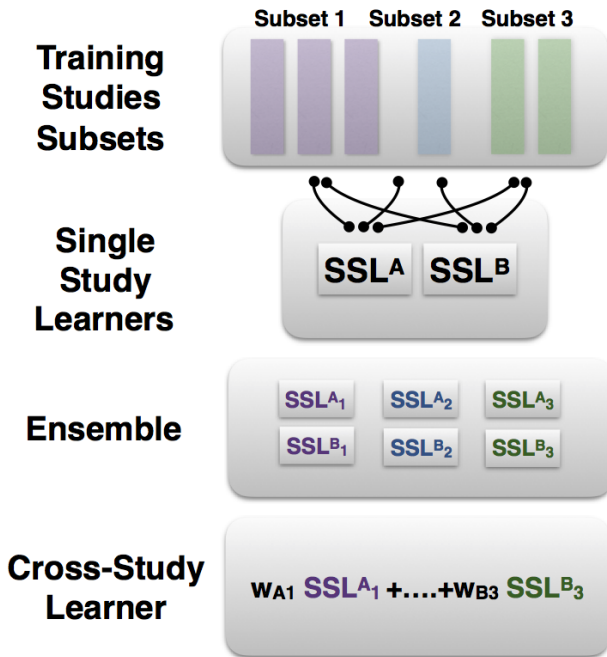




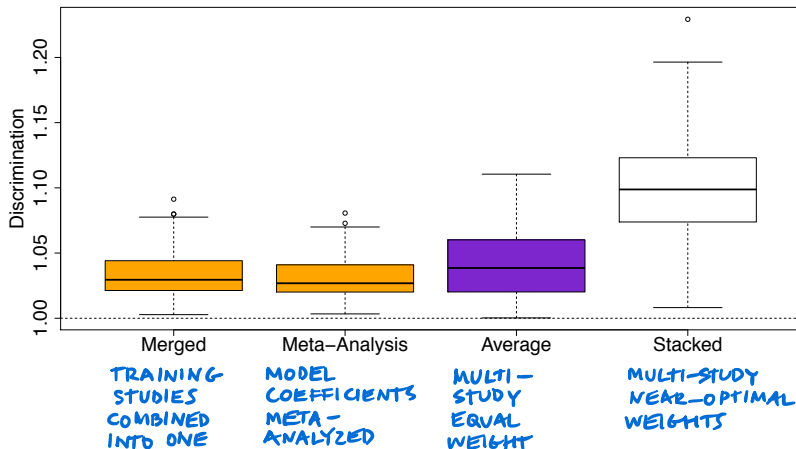
Distribution of ranks



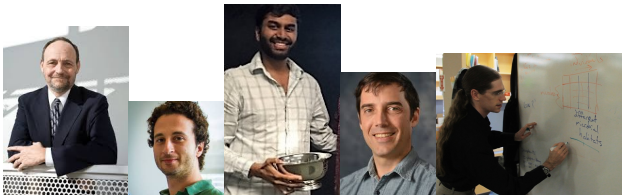
Attention to replicability
requires rethinking existing machine learning principles.



Progress



- 1 Variability across studies is unavoidable; some of this variability genuinely contributes scientific insight.
- 2 To make progress in replicability, it is important to:
 - Work harder at educating the scientific community about how much cross-study variation to expect
 - Work harder at generating curated comprehensive data collections
 - Think harder about existing machine learning and prediction principles.



Ganzfried, B.F., Riester, M., Haibe-Kains, B., Risch, T., Tyekucheva, S., Jazic, I., Wang, X.V., Ahmadifar, M., **Birrer, M.J.**, Parmigiani, G., **Huttenhower, C.**, **Waldron, L.**, 2013. curatedOvarianData: clinically annotated data for the ovarian cancer transcriptome. Database 2013, bat013–bat013.

Waldron, L., Haibe-Kains, B., Culhane, A.C., Riester, M., Ding, J., Wang, X.V., Ahmadifar, M., Tyekucheva, S., Bernau, C., Risch, T., Ganzfried, B.F., Huttenhower, C., Birrer, M., Parmigiani, G., 2014. Comparative meta-analysis of prognostic gene signatures for late-stage ovarian cancer. JNCI 106, dju049–dju049.

Zhao, S.D., Parmigiani, G., Huttenhower, C., Waldron, L., 2014. Más-o-menos: a simple sign averaging method for discrimination in genomic data analysis. Bioinformatics 30, 3062–3069.

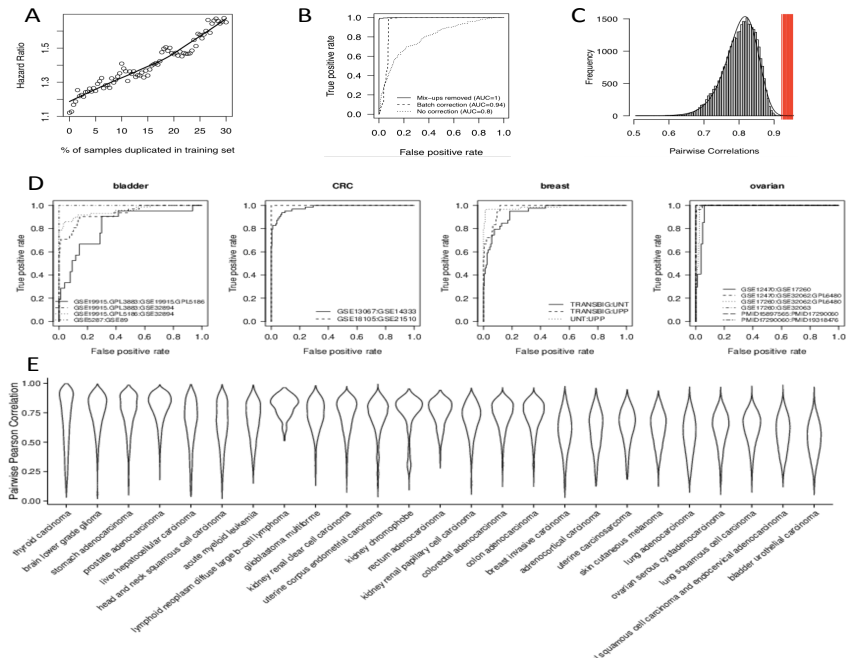
Bernau, C., Riester, M., Boulesteix, A.-L., Parmigiani, G., Huttenhower, C., Waldron, L., **Trippl, L.**, 2014. Cross-study validation for the assessment of prediction algorithms. Bioinformatics 30, i105–12.

Trippl, L., Waldron, L., Huttenhower, C., Parmigiani, G., 2015. Bayesian nonparametric cross-study validation of prediction methods. Ann. Appl. Stat 9, 402–428.

Waldron, L., Riester, M., Ramos, M., Parmigiani, G., Birrer, M., 2016. The Doppelgänger Effect: Hidden Duplicates in Databases of Transcriptome Profiles. JNCI 108, djw146.

Zhang Y, Bernau C, Parmigiani G, Waldron L. The impact of different sources of heterogeneity on loss of accuracy from genomic prediction models. Biostatistics. 2018;6(e1002358):701.

Patil, P. Parmigiani, G. Building replicable predictors via ensembling.

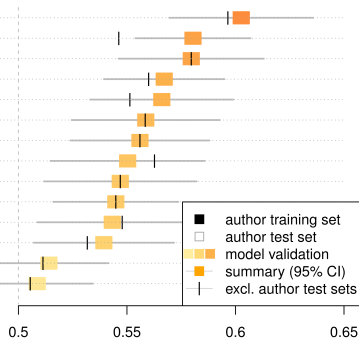


using c-stat instead

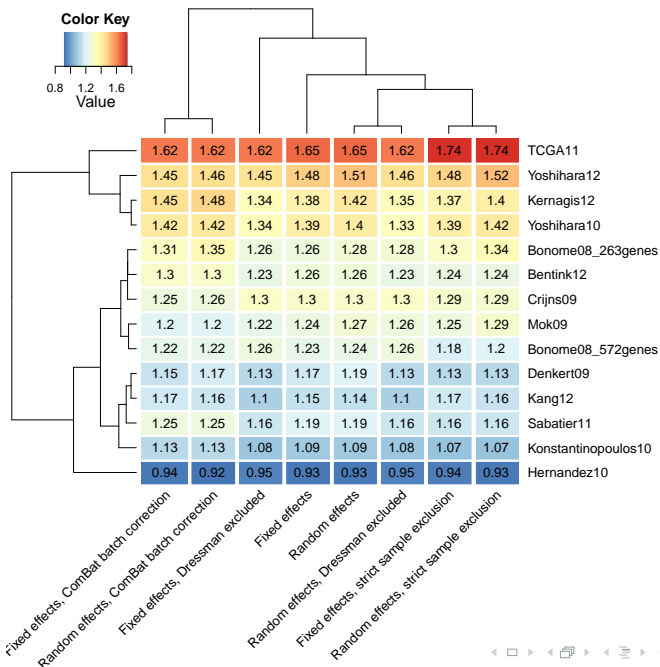
(A) Implemented Models Validation Statistics for 14 Models in 10 Datasets

Dataset Average	0.61	0.58	0.57	0.56	0.56	0.55	0.55	0.54	0.54	0.53
TCGA11	0.62	0.69	0.6	0.63	0.61	0.47	0.57	0.6	0.64	0.55
Yoshihara12	0.63	0.81	0.64	0.6	0.62	0.51	0.5	0.58	0.57	0.55
Bonome08_263genes	0.57	0.68	0.58	0.6	0.62	0.53	0.6	0.54	0.56	0.52
Yoshihara10	0.7	0.55	0.62	0.53	0.55	0.53	0.54	0.8	0.56	0.52
Kernagis12	0.66	0.58	0.63	0.56	0.55	0.55	0.65	0.57	0.55	0.54
Sabatier11	0.64	0.54	0.56	0.57	0.54	0.62	0.55	0.57	0.56	0.52
Crijns09	0.5	0.6	0.59	0.55	0.58	0.55	0.56	0.47	0.54	0.67
Bentink12	0.65	0.56	0.55	0.61	0.55	0.57	0.57	0.53	0.53	0.52
Bonome08_572genes	0.57	0.6	0.54	0.55	0.64	0.63	0.55	0.5	0.53	0.54
Mok09	0.53	0.6	0.56	0.57	0.57	0.53	0.69	0.57	0.51	0.51
Kang12	0.63	0.54	0.52	0.54	0.57	0.54	0.49	0.54	0.58	0.52
Denkert09	0.67	0.52	0.54	0.53	0.53	0.58	0.53	0.51	0.52	0.55
Hernandez10	0.56	0.61	0.56	0.54	0.53	0.5	0.5	0.54	0.49	0.51
Konstantinopoulos10	0.57	0.5	0.52	0.48	0.49	0.6	0.5	0.51	0.53	0.5
Expression Datasets	Dressman	Yoshihara 2012A	Tothill	Bentink	Bonome	Konstantinopoulos	Mok	Yoshihara 2010	TCGA	Crijns

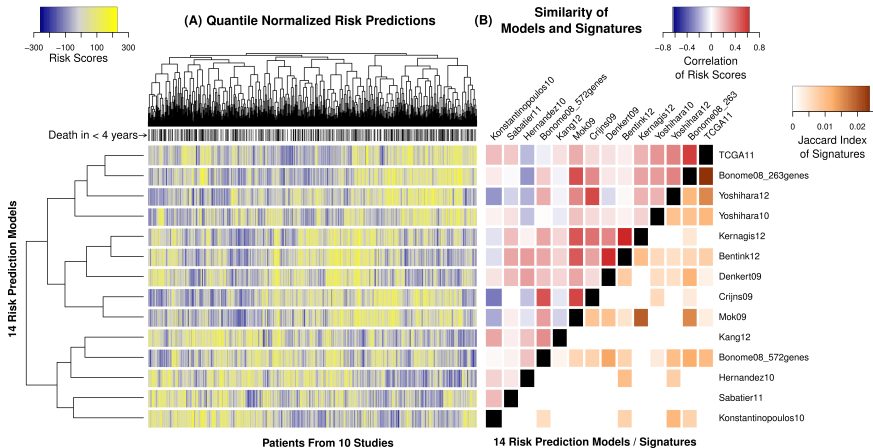
(B) Summary Validation Statistics



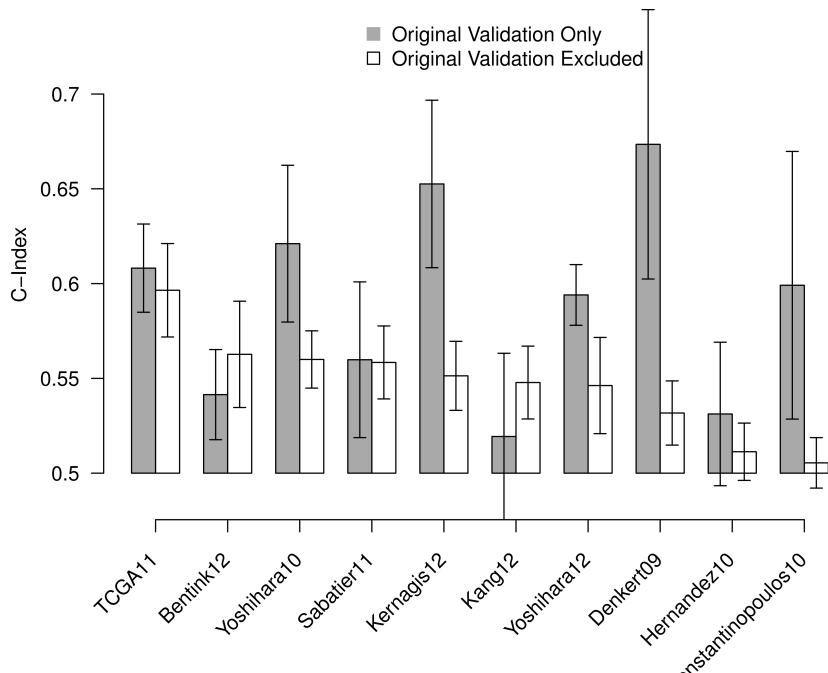
sensitivity analysis



do predictors rank patients similarly?



selection bias in choice of validation study?



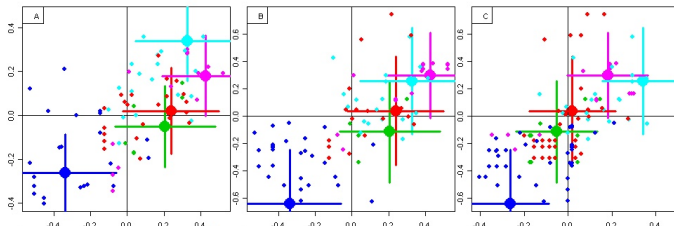


SIGNATURE

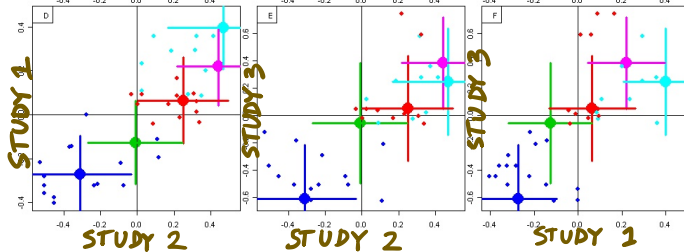


GENES WITHIN
SIGNATURE

UNFILTERED



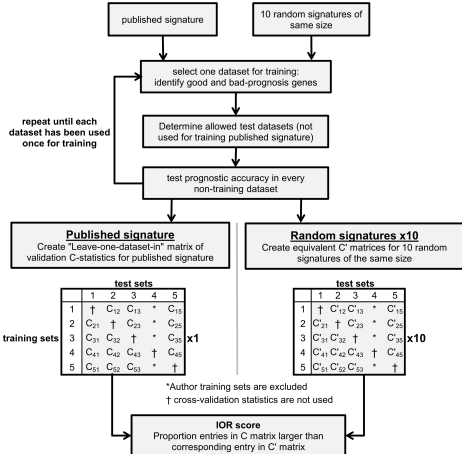
FILTERED



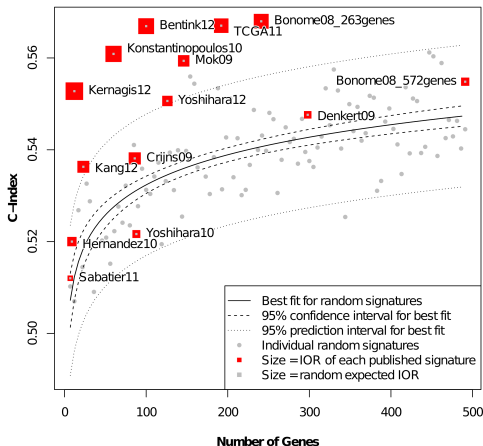
SCALE: PREDICTION STRENGTH (COX EFFECT SIZE)

signatures

A) Methodology for comparing prognostic quality of gene sets to random gene sets



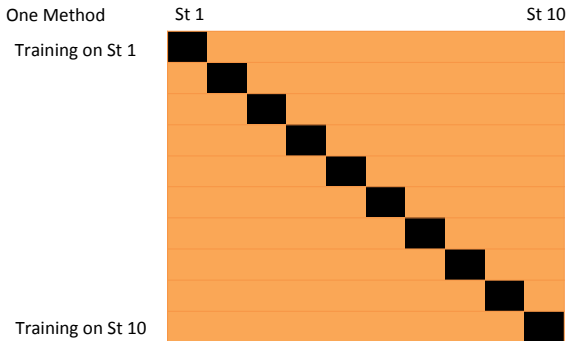
B) Gene set Improvement over Random Signatures



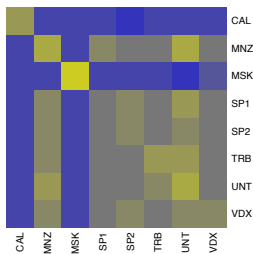
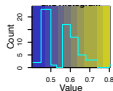
Meta-analysis of prediction performance: evaluates specific classifier

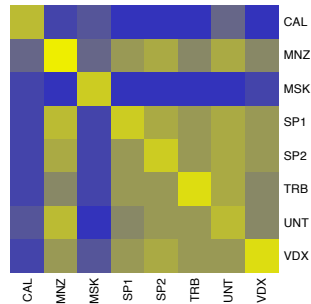
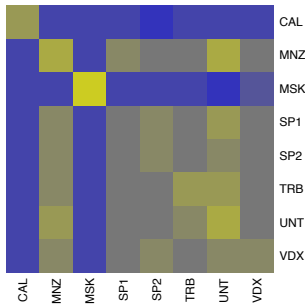
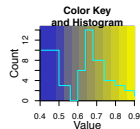
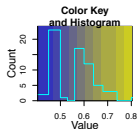


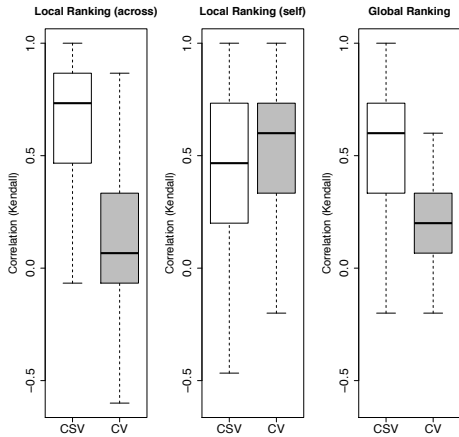
Cross-study performance evaluation: Evaluates Methodology



No.	Name	Adjuvant therapy	# patients	# ER+	3Q survival [mo.]	Median follow-up [mo.]	Original identifiers [‡]	Reference
1	CAL	chemo, hormonal	118	75	42	82	CAL	Chin <i>et al.</i> (2006)
2	MNZ	none	200	162	120	94	MAINZ	Schmidt <i>et al.</i> (2008)
3	MSK	combination	99	57	76	82	MSK	Minn <i>et al.</i> (2005)
4	ST1	hormonal	512*	507*	114	106	MDA5, TAM, VDX3	Foekens <i>et al.</i> (2006)
5	ST2	hormonal	517	325	126	121	EXPO, TAM	Symmans <i>et al.</i> (2010)
6	TRB	none	198	134	143	171	TRANSBIG	Desmedt <i>et al.</i> (2007)
7	UNT	none	133	86	151	105	UNT	Sotiriou <i>et al.</i> (2006)
8	VDX	none	344	209	44	107	VDX	Minn <i>et al.</i> (2007)



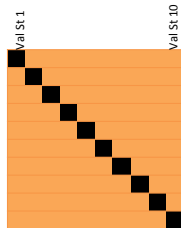




Cross-study performance evaluation: Evaluates Methodology

One Method

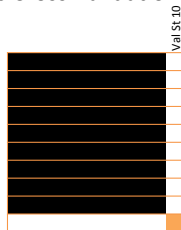
Train on St 1



Train on St 10

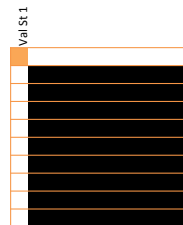
Meta-analytic Cross Validation: Train Classifier using all studies

Train on St 1--9



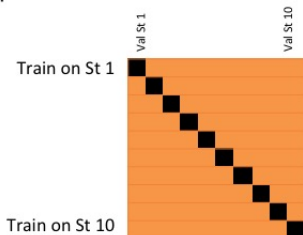
.....

Train on St 2--10



training on many studies

Cross-study performance evaluation: Evaluates Methodology



Cross-study Training for Replicability: Trains new Predictors

