

Scoring the Multiple Sclerosis Diagnosis Sub-Challenge

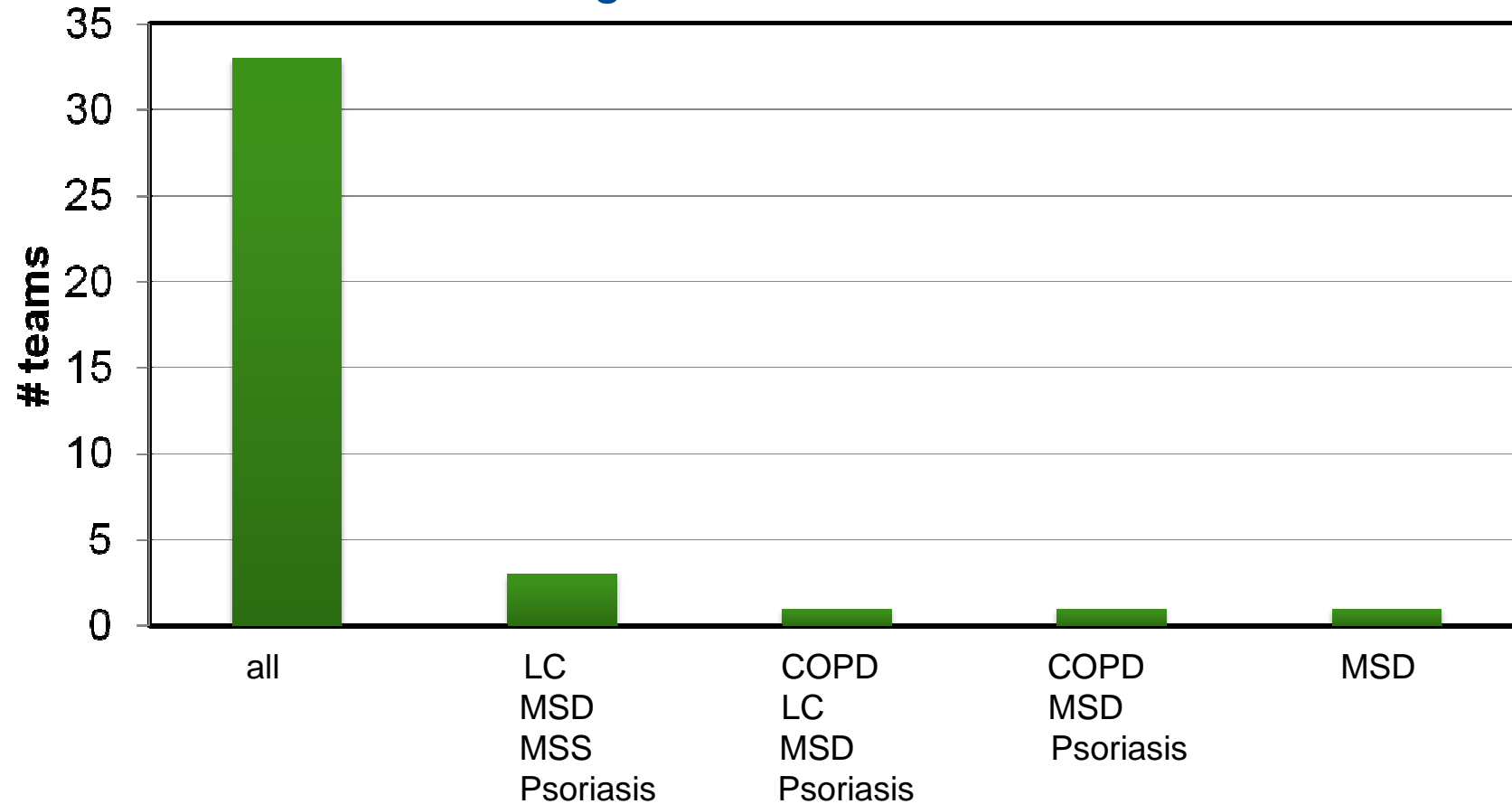
Erhan Bilal, Ph.D.,
Functional Genomics & Systems Biology,
IBM Computational Biology Center
2nd October 2012



Outline

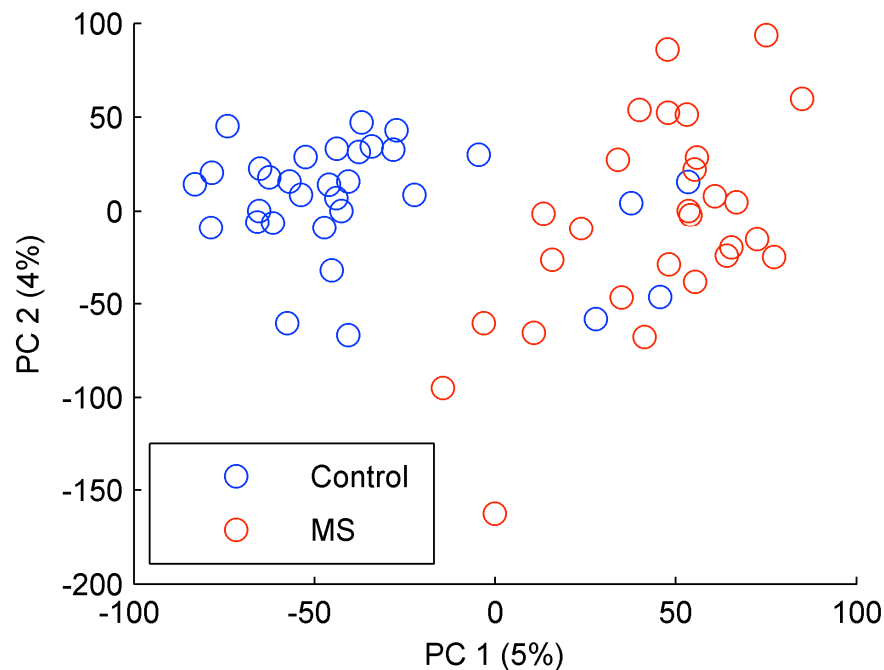
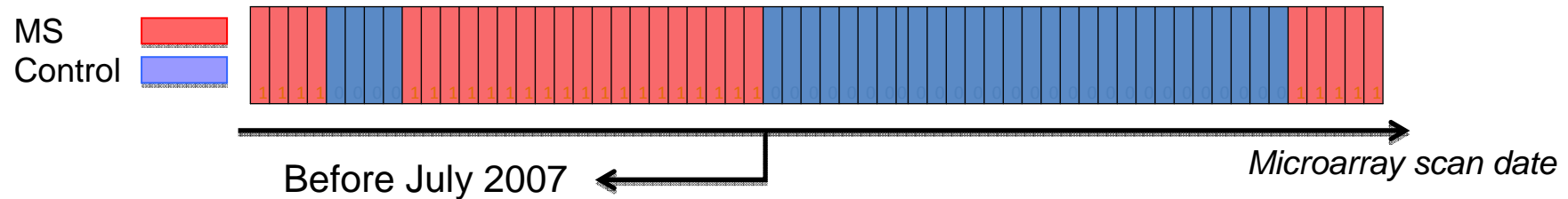
- Participation
- Lessons learned on experimental design and challenge formulation
- Overview of predictions
- Comparisons with random predictions
- Robustness of Ranking
- Final results

Most teams that participated in the MS challenge also participated in all the other sub-challenges



The challenge prize was to be awarded to the teams that participated in all challenges. This encouragement to participate in all challenges proved effective.

Lessons about experimental design and challenge formulation



Unsupervised clustering of the test set using PCA

Two clusters emerge in an unsupervised PCA analysis of the test set.

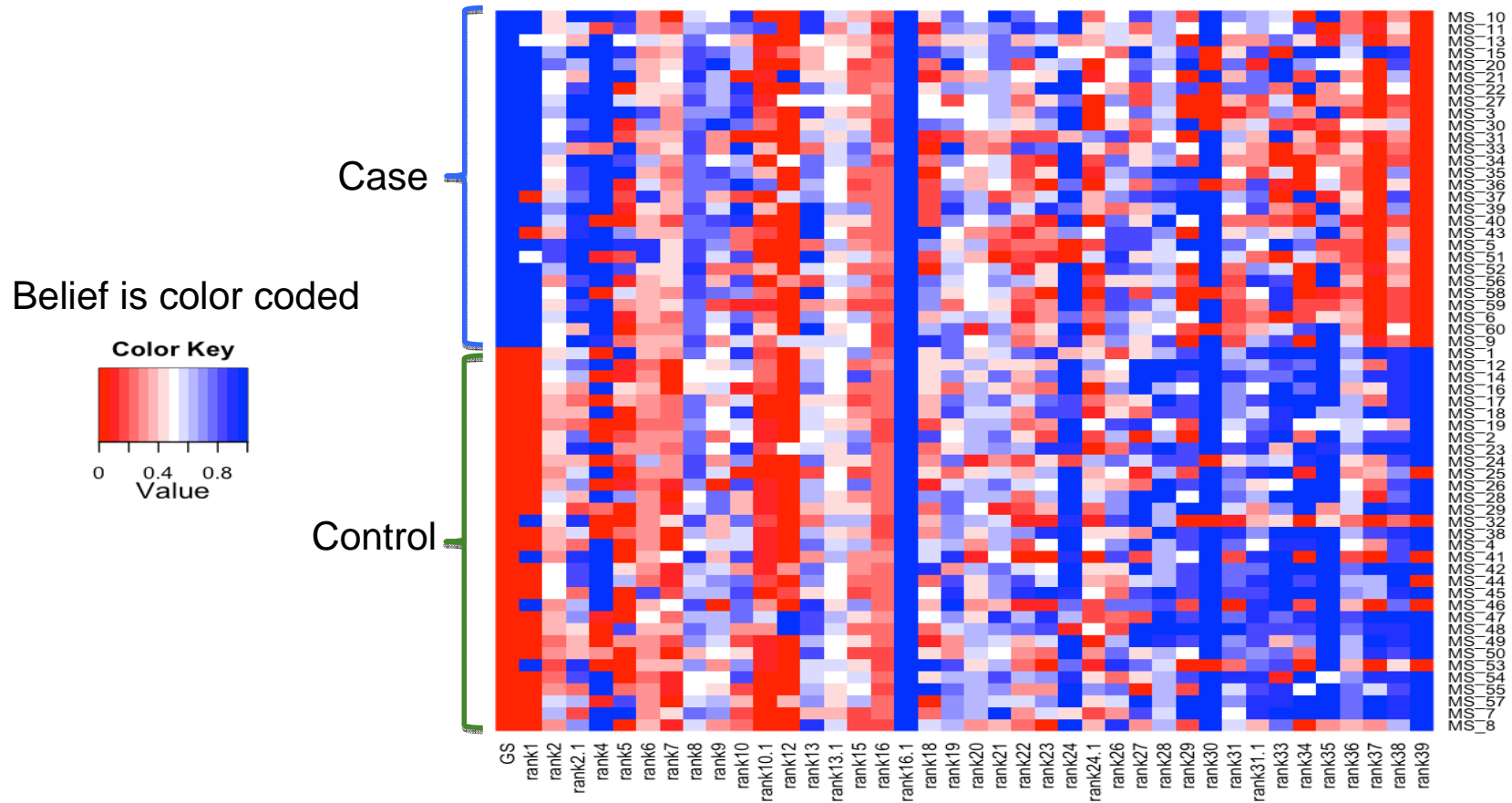
Phenotype and microarray scan date are highly correlated: The clusters could have originated from batch effects or from the phenotype.

The original experiments would have been better designed if cases and controls had been uncorrelated with respect to the dates.

Lesson 1: Giving away the test set can give unintended information to participants.

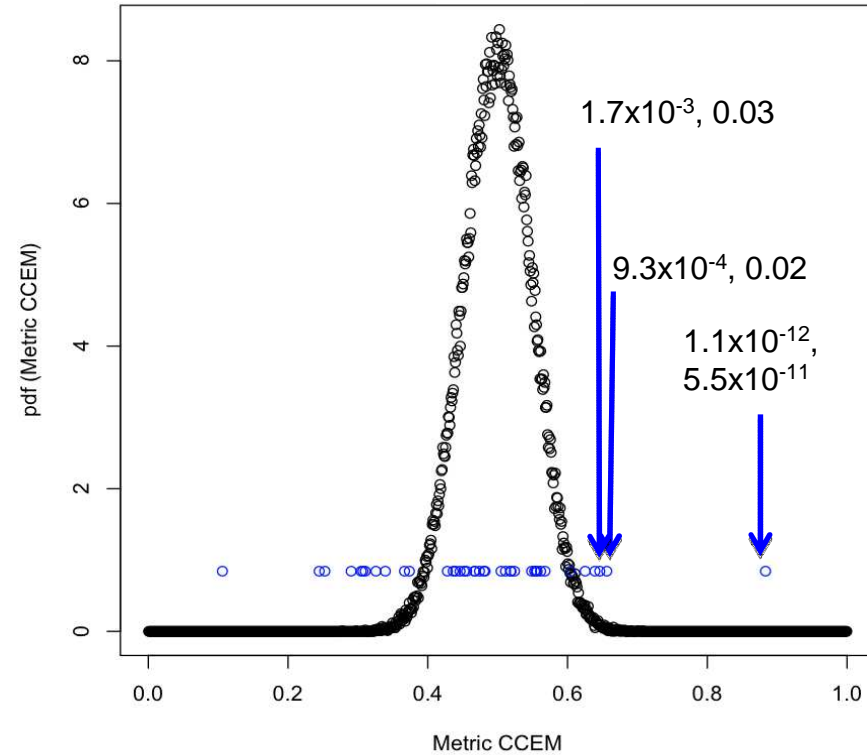
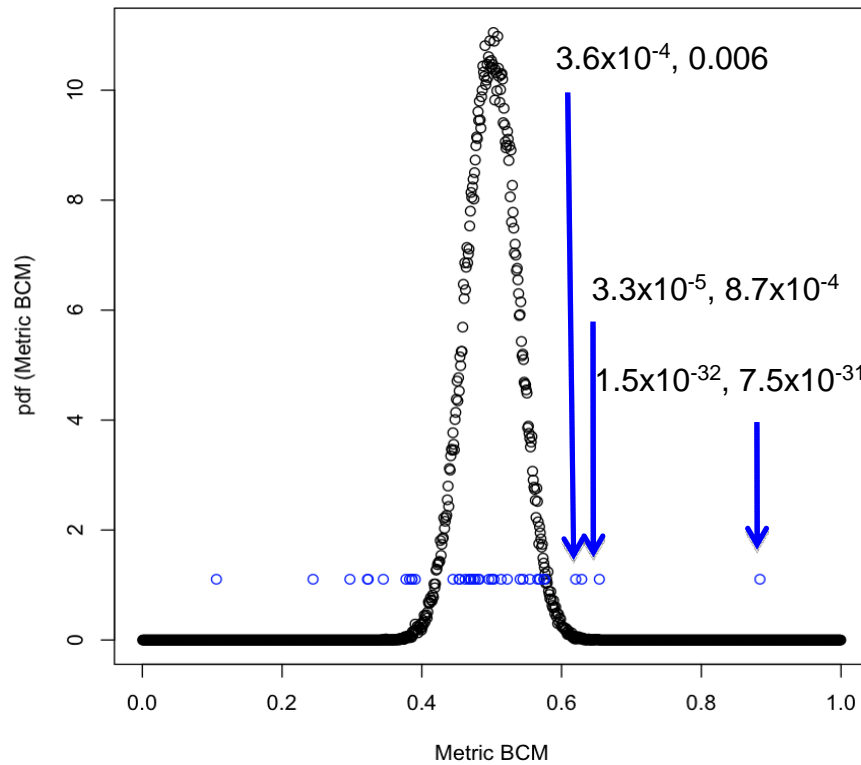
Lesson 2: Next iteration of this challenge should test methods one patient at a time (patients go to the doctor without their cohorts!)

Overview of predictions



- There are a number of teams that are almost perfectly anti-correlated with the GS. Is this a byproduct of the batch clusters?
- Overall the quality of the predictions looks worse than in the other sub-challenges.
- No obvious systematic patient misclassification was observed.

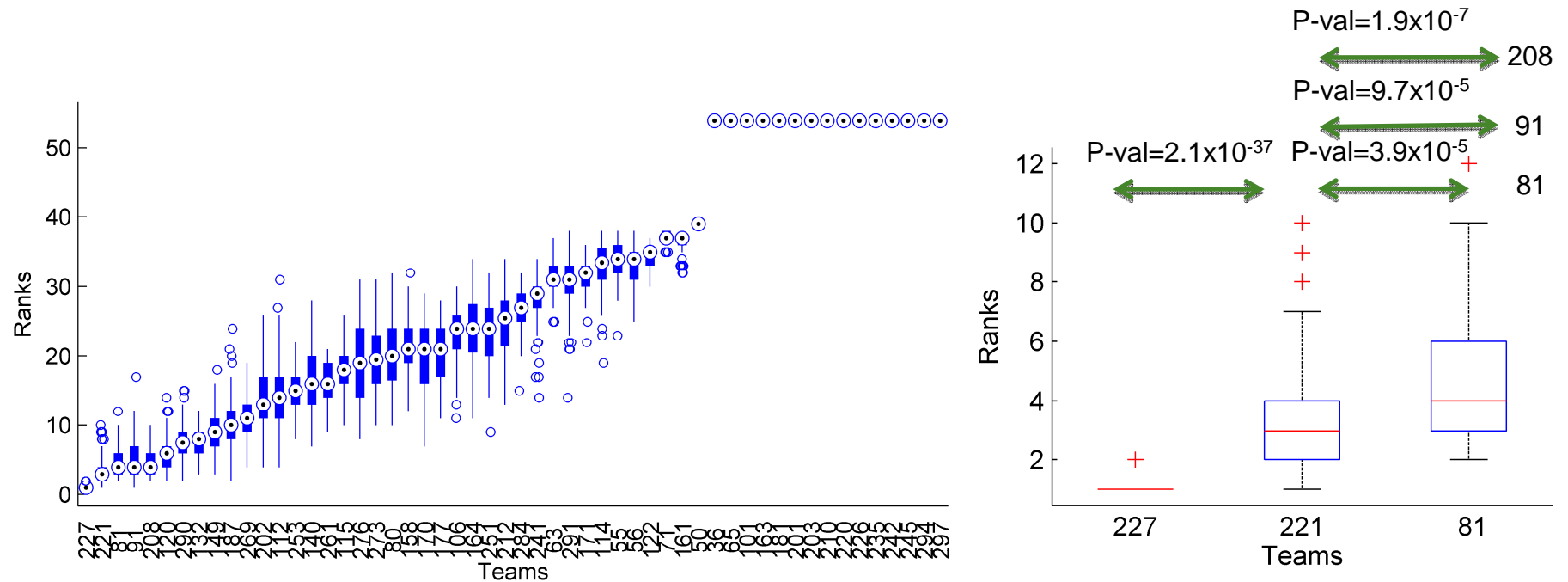
Comparison of performance with random predictions: The “absolute” null hypothesis



- Some teams did much worse than random. This is probably due to mislabeling the clusters in the test set.
- There are some teams that did extremely well, and picked the real signal.

Robustness of ranking: bootstrapping the gold-standard

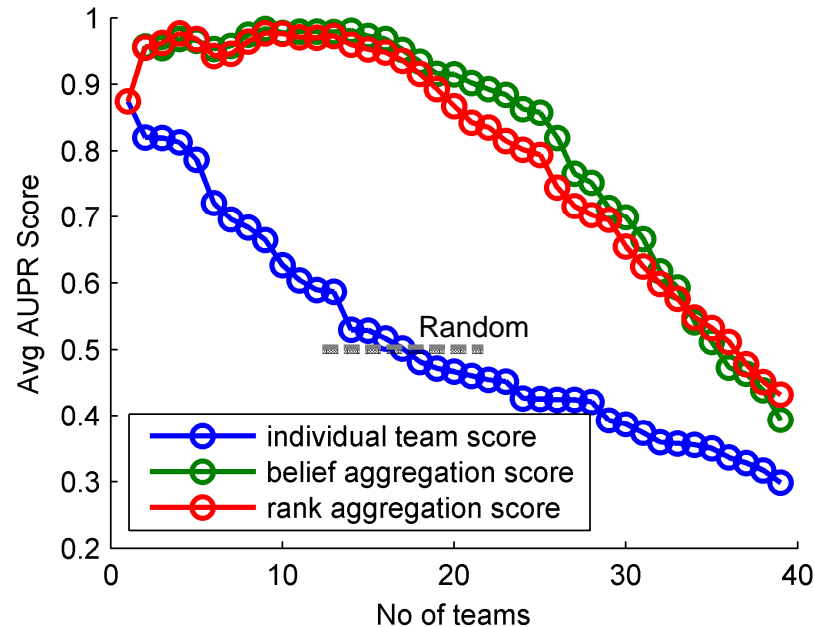
P-values were computed using the rank-sum Wilcoxon test for the ranks obtained by the teams in different bootstrapping realizations of the gold-standard.



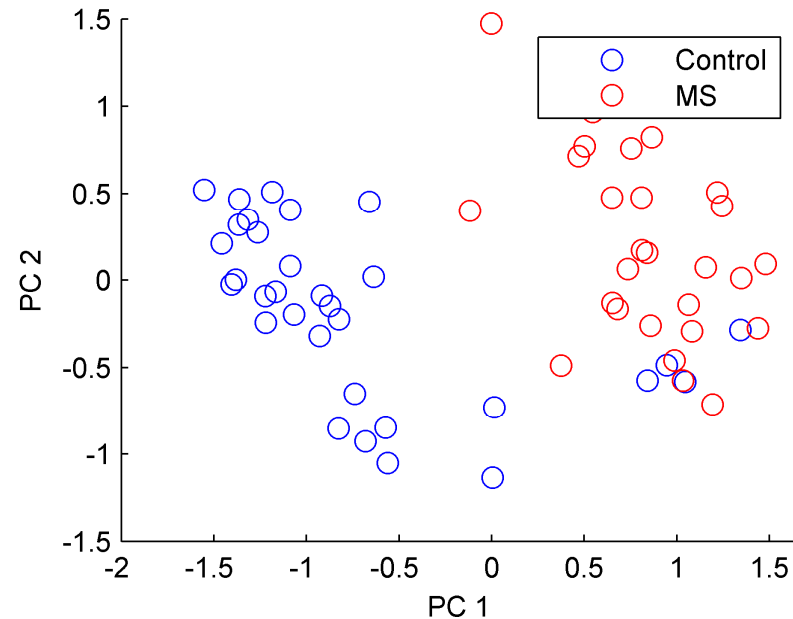
- The ranking of the top performers is robust to the composition of the Gold Standard.
- The ranks of the participants with medium performance depend more on the specific composition of the Gold Standard.

Aggregation of results: The wisdom of crowds

Average rank and average belief aggregation



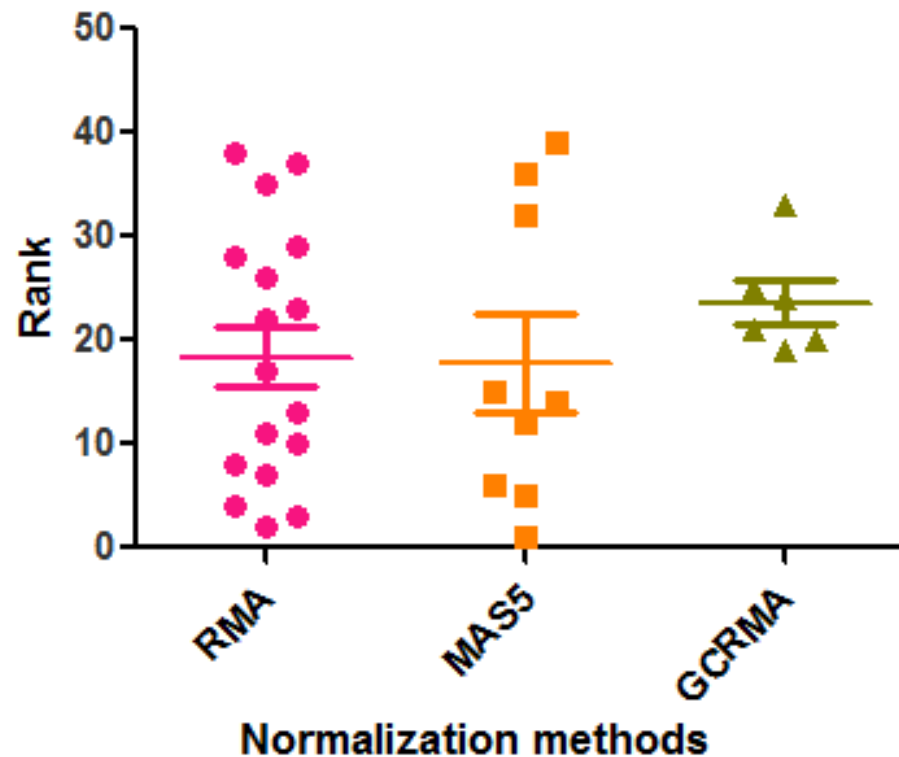
Unsupervised PCA analysis of teams' predictions



Aggregation of top performers improves on the prediction of the best performer in this challenge. Adding worse than random teams degrades aggregation.

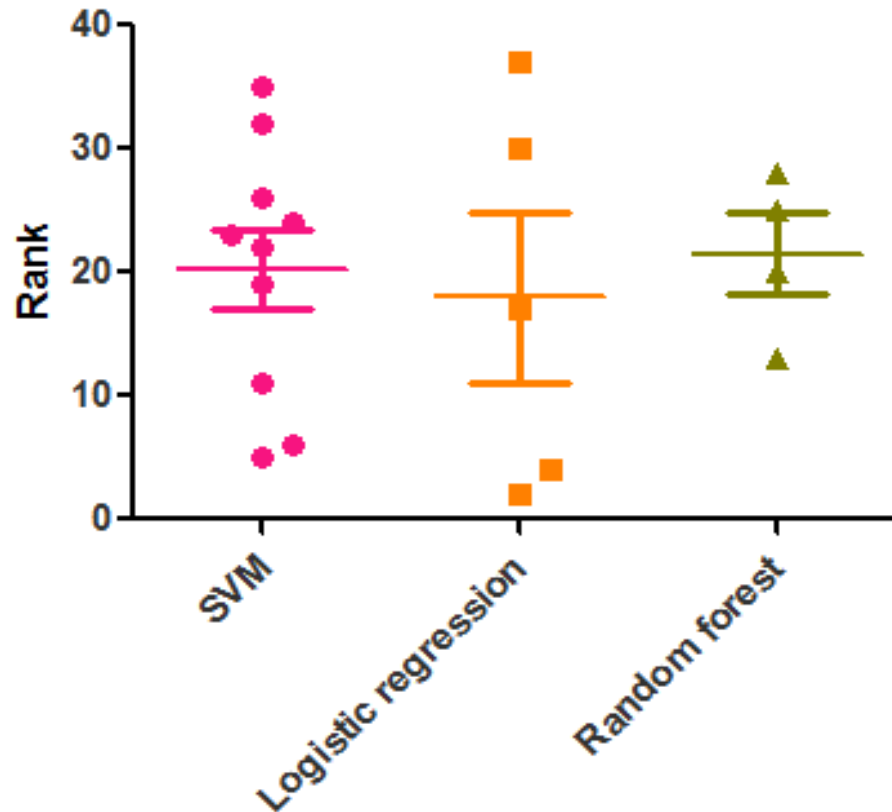
The collective predictions have enough information to cluster the subjects almost perfectly in cases and controls.

No normalization method conferred an added advantage



- Participants used the typical normalization methods such as MAS5.0 (28%) , RMA (53%) and GCRMA (19%).
- No advantage was conferred by the use of any of these normalization methods.

No classification method conferred an added advantage



- The most used classification methods after feature selection were SVM, Logistic Regression and Random Forest methods.
- No specific advantage was conferred by the use of any of these classification methods.

Final ranking

Team	BCM	CCEM	AUPR_avg	Rank-sum	Rank
Team227	0.88	0.88	0.87	3	1
Team221	0.63	0.63	0.82	11	2
Team208	0.58	0.64	0.82	11	2
Team91	0.65	0.65	0.70	12	4
Team81	0.62	0.66	0.68	14	5
Team120	0.57	0.60	0.81	17	6
Team290	0.57	0.56	0.72	20	7
Team132	0.57	0.55	0.79	23	8
Team149	0.55	0.57	0.67	25	9
Team187	0.54	0.55	0.63	31	10
Team269	0.54	0.56	0.60	31	10
Team202	0.51	0.55	0.53	40	12
Team112	0.52	0.48	0.59	43	13
Team253	0.50	0.52	0.53	43	13
Team140	0.50	0.51	0.52	47	15
Team261	0.50	0.47	0.59	50	16
Team115	0.50	0.52	0.47	50	16
Team276	0.48	0.52	0.45	54	18
Team80	0.48	0.47	0.47	58	19
Team158	0.48	0.50	0.43	61	20

Team	BCM	CCEM	AUPR_avg	Rank-sum	Rank
Team177	0.47	0.47	0.46	63	21
Team273	0.47	0.46	0.48	64	22
Team170	0.47	0.44	0.50	65	23
Team164	0.47	0.45	0.43	73	24
Team251	0.45	0.48	0.42	73	24
Team106	0.46	0.45	0.42	76	26
Team212	0.44	0.43	0.45	79	27
Team284	0.45	0.44	0.39	84	28
Team241	0.39	0.37	0.42	85	29
Team291	0.39	0.37	0.39	89	30
Team63	0.39	0.34	0.35	97	31
Team171	0.38	0.31	0.37	97	31
Team114	0.34	0.33	0.36	99	33
Team56	0.32	0.31	0.36	101	34
Team55	0.32	0.30	0.36	104	35
Team122	0.38	0.29	0.33	106	36
Team161	0.24	0.25	0.34	111	37
Team71	0.30	0.24	0.32	113	38
Team50	0.11	0.11	0.30	117	39

Best performing team in the MS Diagnostic IMPROVER Sub-Challenge

Team	BCM	CCEM	AUPR_avg	Rank-sum	Rank
Team227	0.88	0.88	0.87	3	1

Team: COSBI

Team Leader: Mario Lauria

Institution: COSBI (Computational Systems Biology), Italy

Congratulations!