CAID

June 24, 2019

Contents

Chapter 1

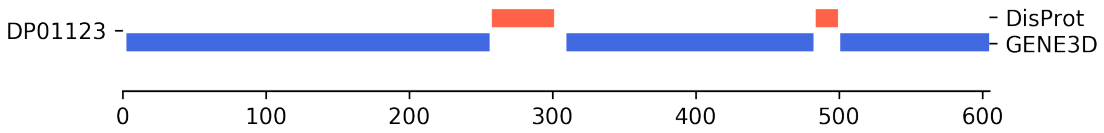
CAID

1.1 Dataset

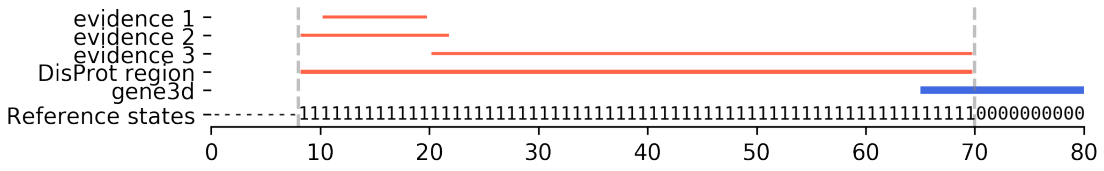
Critical Assessment of Intrinsic Disorder (CAID) is a continuous experiment where prediction methods for intrinsic disorder (ID) are blind tested on unpublished DisProt data.

```
Accordion(children=(RadioButtons(description='Reference:', index=1, options=('new-pdb-missing', 'new-dis
```

Current analysis is performed on the **new-disprot-all** dataset with **gene3d** negative definition. This means that DisProt defines order for *all its new entries* and *gene3d defines order*.



DisProt entries can have annotation covering the same sequence space. In these cases different evidences are merged in a unique continuous region. DisProt merged regions define positive cases (labeled as 1) in reference states. Remaining states are considered undefined.



1.2 Evaluation metrics

Metrics evaluating prediction scores are calculated applying **Default** thresholds to prediction scores. Table is sorted by descending value of BAc column

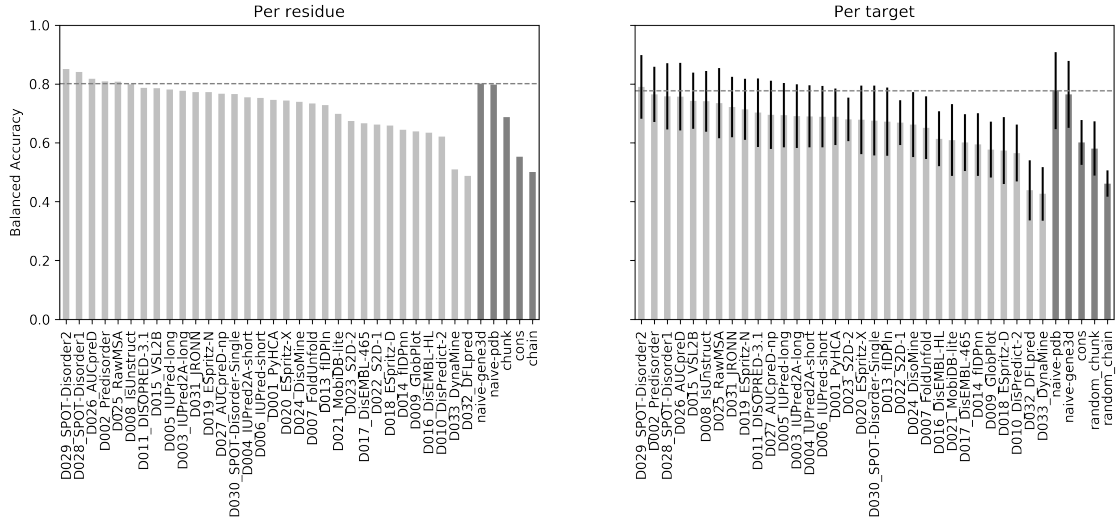
	FN	FP	TN	TP	BAc	F1s	MCC	Pre	Rec	Rec_n	AUC_PRC	AUC_ROC	npred	nref
D029_SPOT-Disorder2	11513	7577	124439	36238	0.851	0.792	0.722	0.827	0.759	0.943	0.863	0.924	610	646
D028_SPOT-Disorder1	13693	10500	144865	40600	0.840	0.770	0.694	0.795	0.748	0.932	0.828	0.923	644	646
D026_AUCpreD	17070	8046	144327	37485	0.817	0.749	0.675	0.823	0.687	0.947	0.836	0.913	644	646
D002_Predisorder	10526	28820	121902	43971	0.808	0.691	0.568	0.604	0.807	0.809	0.795	0.885	642	646
D025_RawMSA	17363	10702	144954	37241	0.807	0.726	0.641	0.777	0.682	0.931	0.794	0.899	646	646
naive-gene3d	6767	42457	112552	47544	0.801	0.659	0.533	0.528	0.875	0.726	0.714	0.809	639	646
D008_IsUnstruct	13771	23195	132461	40833	0.799	0.688	0.571	0.638	0.748	0.851	0.755	0.874	646	646
naive-pdb	6072	45103	109201	48119	0.798	0.653	0.525	0.516	0.888	0.708	0.696	0.804	630	646
D011_DISOPRED-3.1	19395	11288	144368	35209	0.786	0.697	0.605	0.757	0.645	0.927	0.771	0.873	646	646
D015_VSL2B	10059	37996	117369	44234	0.785	0.648	0.512	0.538	0.815	0.755	0.758	0.867	644	646
D005_IUPred-long	19362	12932	142652	35198	0.781	0.686	0.586	0.731	0.645	0.917	0.762	0.865	645	646
D003_IUPred2A-long	19974	12452	143204	34630	0.777	0.681	0.583	0.736	0.634	0.920	0.758	0.863	646	646
D031_JRONN	14112	30608	124976	40448	0.772	0.644	0.505	0.569	0.741	0.803	0.725	0.845	645	646
D019_ESpritz-N	17054	22388	133196	37506	0.772	0.655	0.528	0.626	0.687	0.856	0.722	0.839	645	646
D027_AUCpreD-np	23312	6080	149576	31292	0.767	0.680	0.612	0.837	0.573	0.961	0.797	0.888	646	646
D030_SPOT-Disorder-Single	24220	3796	151860	30384	0.766	0.684	0.632	0.889	0.556	0.976	0.823	0.902	646	646
D004_IUPred2A-short	23880	8535	147121	30724	0.754	0.655	0.571	0.783	0.563	0.945	0.740	0.863	646	646
D006_IUPred-short	23891	8930	146654	30669	0.752	0.651	0.566	0.774	0.562	0.943	0.736	0.861	645	646
D001_PyHCA	18648	25851	129805	35956	0.746	0.618	0.474	0.582	0.658	0.834	0.661	0.820	646	646
D020_ESpritz-X	25551	7237	148347	29009	0.743	0.639	0.563	0.800	0.532	0.953	0.747	0.867	645	646
D024_DisoMine	23260	14920	140736	31344	0.739	0.621	0.506	0.678	0.574	0.904	0.664	0.861	646	646
D007_FoldUnfold	17940	30485	122269	35715	0.733	0.596	0.438	0.540	0.666	0.800	NaN	NaN	621	646
D013_fIDPln	27013	7675	147981	27505	0.728	0.613	0.534	0.782	0.505	0.951	0.733	0.877	645	646
D021_MobiDB-lite	32125	2309	153275	22435	0.698	0.566	0.539	0.907	0.411	0.985	0.752	0.848	645	646
chunk	25255	25255	130400	29348	0.688	0.537	0.375	0.537	0.537	0.838	NaN	NaN	100	100
D023_S2D-2	7710	79130	76235	46583	0.674	0.518	0.312	0.371	0.858	0.491	0.476	0.762	644	646
D017_DisEMBL-465	33266	8642	146723	21027	0.666	0.501	0.417	0.709	0.387	0.944	0.608	0.778	644	646
D022_S2D-1	7237	84331	71034	47056	0.662	0.507	0.293	0.358	0.867	0.457	0.662	0.805	644	646
D018_ESpritz-D	35346	5631	149953	19214	0.658	0.484	0.429	0.773	0.352	0.964	0.682	0.870	645	646
D014_fIDPnn	38167	2000	153656	16351	0.644	0.449	0.446	0.891	0.300	0.987	0.764	0.888	645	646
D009_GlobPlot	33642	16473	139111	20918	0.639	0.455	0.318	0.559	0.383	0.894	0.512	0.708	645	646
D016_DisEMBL-HL	25527	40580	114785	28766	0.634	0.465	0.250	0.415	0.530	0.739	0.480	0.702	644	646
D010_DisPredict-2	31912	27054	128602	22692	0.621	0.435	0.249	0.456	0.416	0.826	0.450	0.689	646	646
cons	4676	126208	29448	49928	0.552	0.433	0.123	0.283	0.914	0.189	0.395	0.666	646	646
D033_DynaMine	53609	64	155520	951	0.509	0.034	0.108	0.937	0.017	1.000	0.684	0.828	645	646
chain	40398	40398	115257	14205	0.500	0.260	0.001	0.260	0.260	0.740	NaN	NaN	100	100
D032_DFLpred	49496	18506	137150	5108	0.487	0.131	-0.035	0.216	0.094	0.881	0.219	0.367	646	646

Where table column names mean:

label	meaning
BAc	balanced accuracy
F1s	F1-score
MCC	Matthew's Correlation Coefficient
Pre	Precision/Selectivity
Rec	Recall/Sensitivity
Rec_n	Specificity
AUC_ROC	Area under the ROC curve
AUC_PRC	Area under the PR curve
npred	number of predicted targets
nref	number of targets in reference

1.2.1 Balanced accuracy

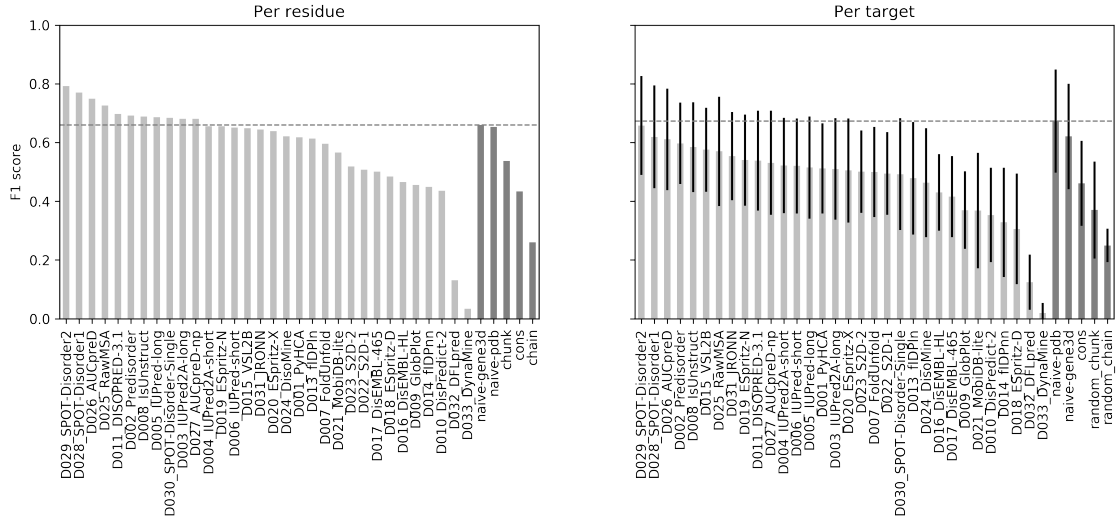
Comparison of predictors and baselines performance by balanced accuracy.



Overall (left panel) and average per-target (right panel) balanced accuracy. Light gray bars represent published prediction methods; dark gray bars represent baseline prediction methods. On the right panel standard deviation is shown as an error bar.

1.2.2 F1-score

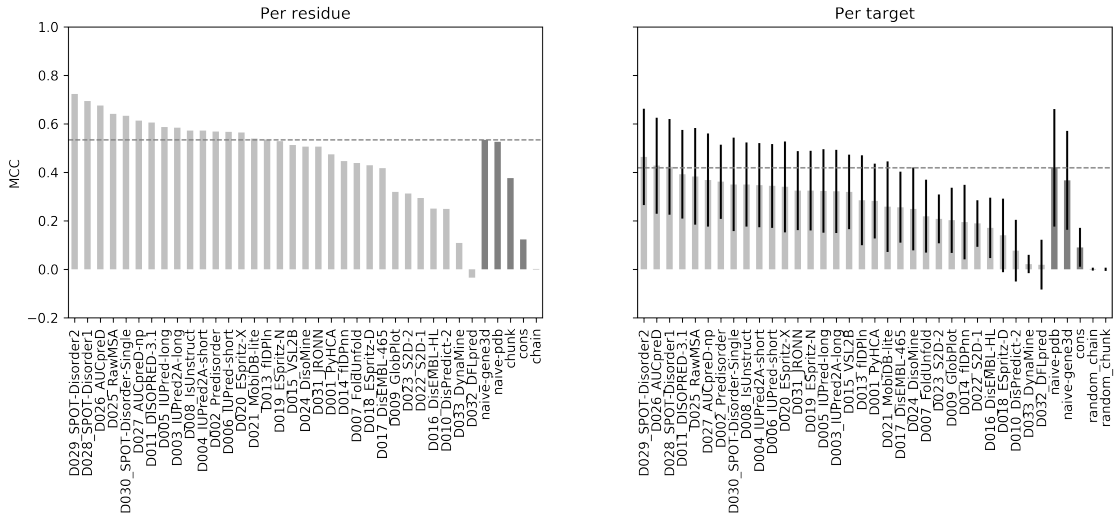
Comparison of predictors and baselines performance by F1-score.



Overall (left panel) and average per-target (right panel) F1-score. Light gray bars represent published prediction methods; dark gray bars represent baseline prediction methods. On the right panel standard deviation is shown as an error bar.

1.2.3 MCC

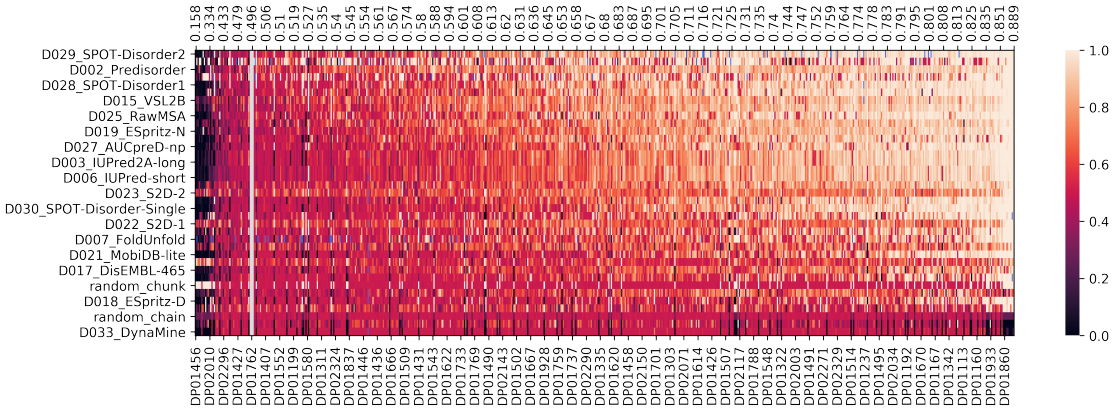
Comparison of predictors and baselines performance by Matthew's Correlation Coefficient.



Overall (left panel) and average per-target (right panel) MCC. Light gray bars represent published prediction methods; dark gray bars represent baseline prediction methods. On the right panel standard deviation is shown as an error bar.

1.2.4 Per target accuracy

Balanced accuracy score for each target for each prediction methods (including baslines)



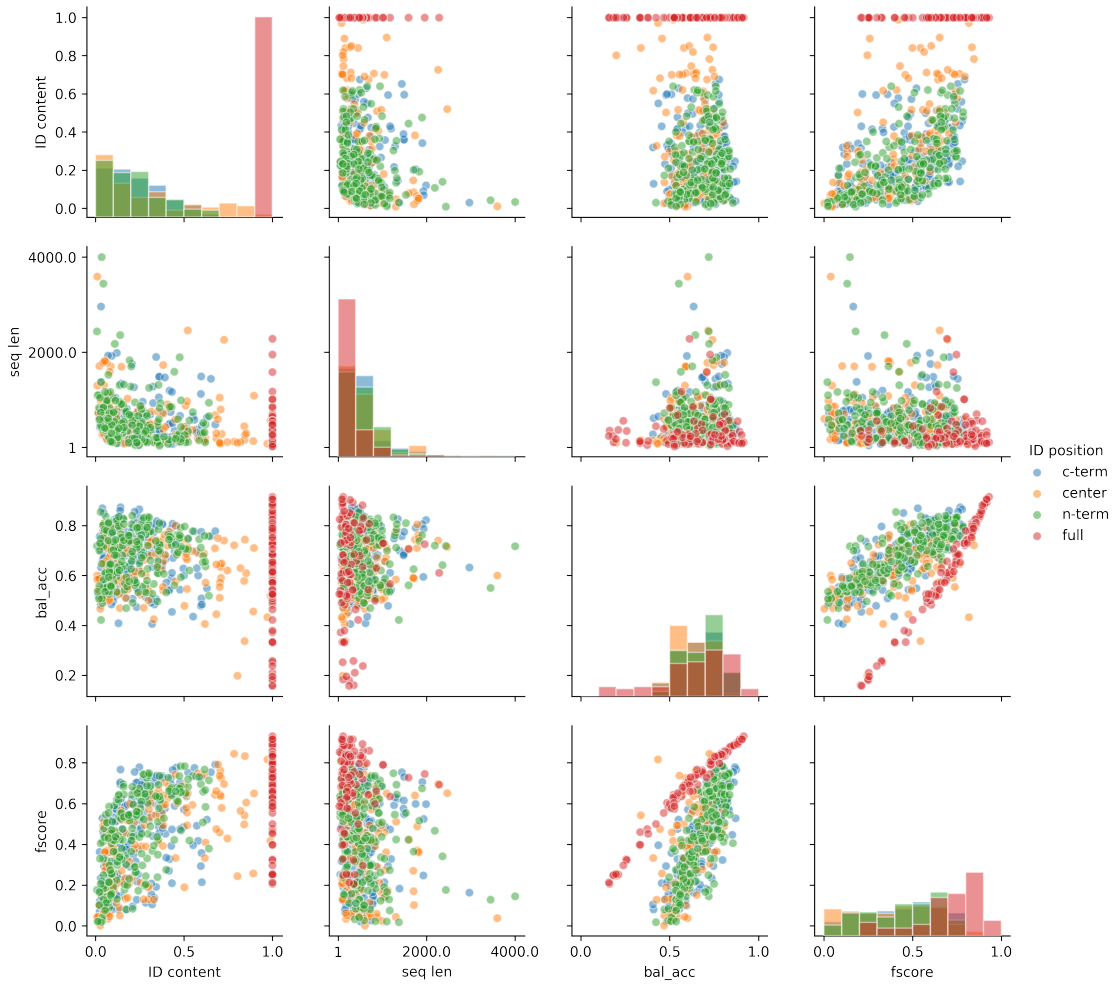
Heatmap of the predictors accuracy for each target. The higher the accuracy the lighter the color. Non-predicted targets are shown in blue. x and y axes are sorted by average accuracy over rows and columns respectively. A white semi-transparent vertical line marks the point where the average accuracy scores for a target is below (left) or above (right) 0.5. Accuracy score approaches 0.5 for a random classifier. Accuracy < 0.5 indicates anti-correlation between predicted and reference classes. Targets with an average accuracy score < 0.5 are:

DP01456	DP01196	DP01248	DP01432	DP01971	DP01870
DP01366	DP01949	DP01281	DP01339	DP01177	DP02010
DP01181	DP01128	DP01512	DP02149	DP01501	DP01498
DP01278	DP01806	DP01898	DP01883	DP02296	DP01584

DP01307	DP01499	DP01285	DP01287	DP02168	DP01195
DP01505	DP01600	DP01430	DP01427	DP01494	DP01907
DP01140	DP01612	DP01500	DP02234	DP01355	DP02025
DP01145	DP01503	DP01762	DP01967	DP02169	DP01187
DP01288					

1.2.5 Target correlation matrix

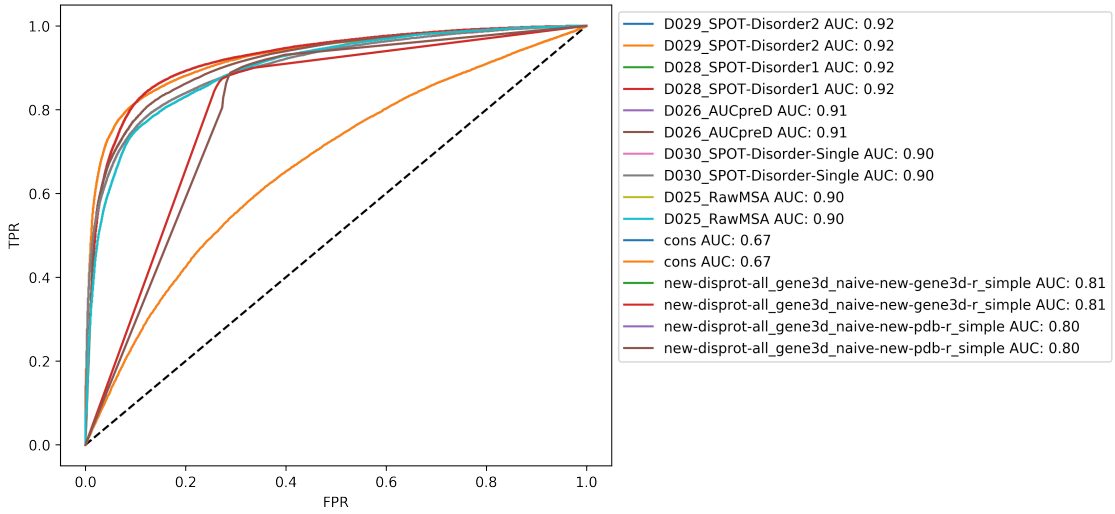
Commonly, experimental data has a bias for low disorder-content. DisProt targets have high disorder-content. Classifiers have been trained/engineered on low disorder-content. I expect difficult targets to have high disorder content. To verify is there is any correlation between target features I'm plotting 4 selected features against each other (Balanced accuracy, F1-Score, Sequence length and ID content). A fifth feature (ID position) divides the datasets in subsets. ID position is calculated as the average of the indexes of disordered residues along the sequence. A correlation is observed in a subset if its points gather around a diagonal.



Correlation matrix of Balanced accuracy, F1-Score, Sequence length and ID content. Average position of disorder is color-coded. Figure matrix is symmetrical. Plots along diagonal axis display single variables distributions. No meaningful correlation is observed.

1.2.6 ROC curve

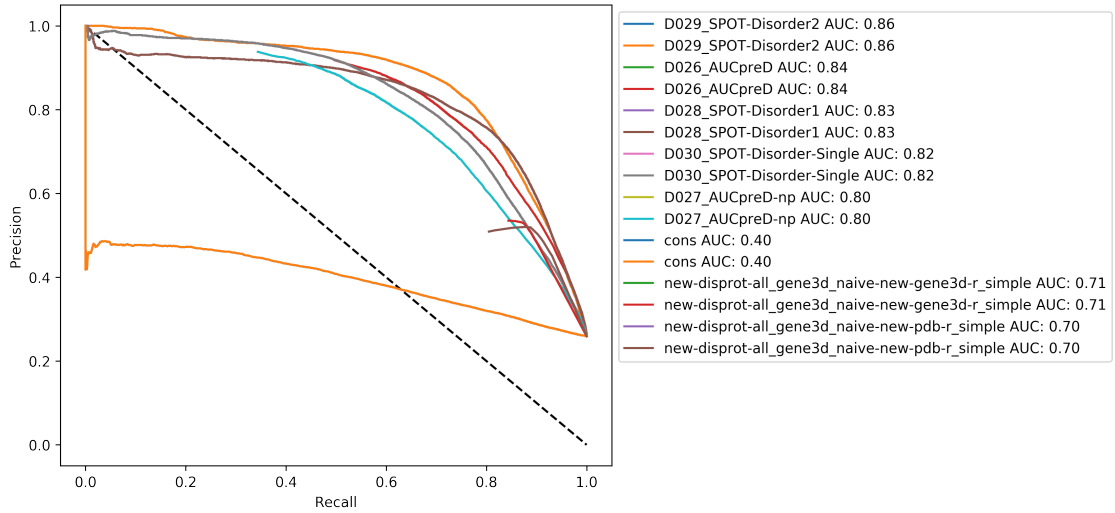
Receiver Operator Characteristic plot for predictors and baselines



False Positive Rate (FPR) on x axis, True Positive Rate (TPR) on y axis. Methods are sorted by their Area Under the Curve (AUC). Only first ten methods plus baselines are shown.

1.2.7 PR curve

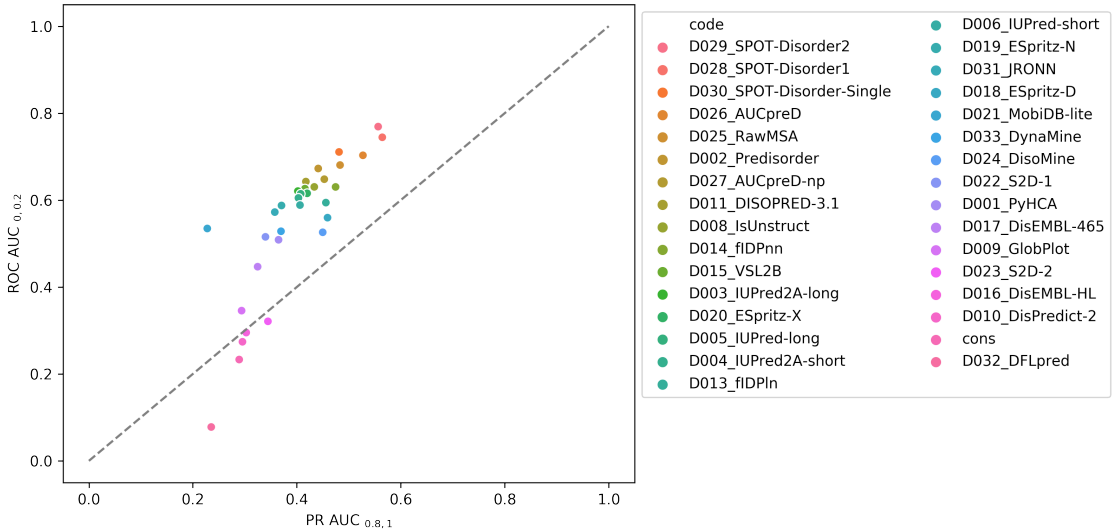
Precision Recall curve plot for predictors and baselines



Recall/Sensitivity on x axis, Precision/Selectivity on y axis. Methods are sorted by their Area Under the Curve (AUC). Only first ten methods plus baselines are shown.

1.2.8 pROC/pPR scatter plot

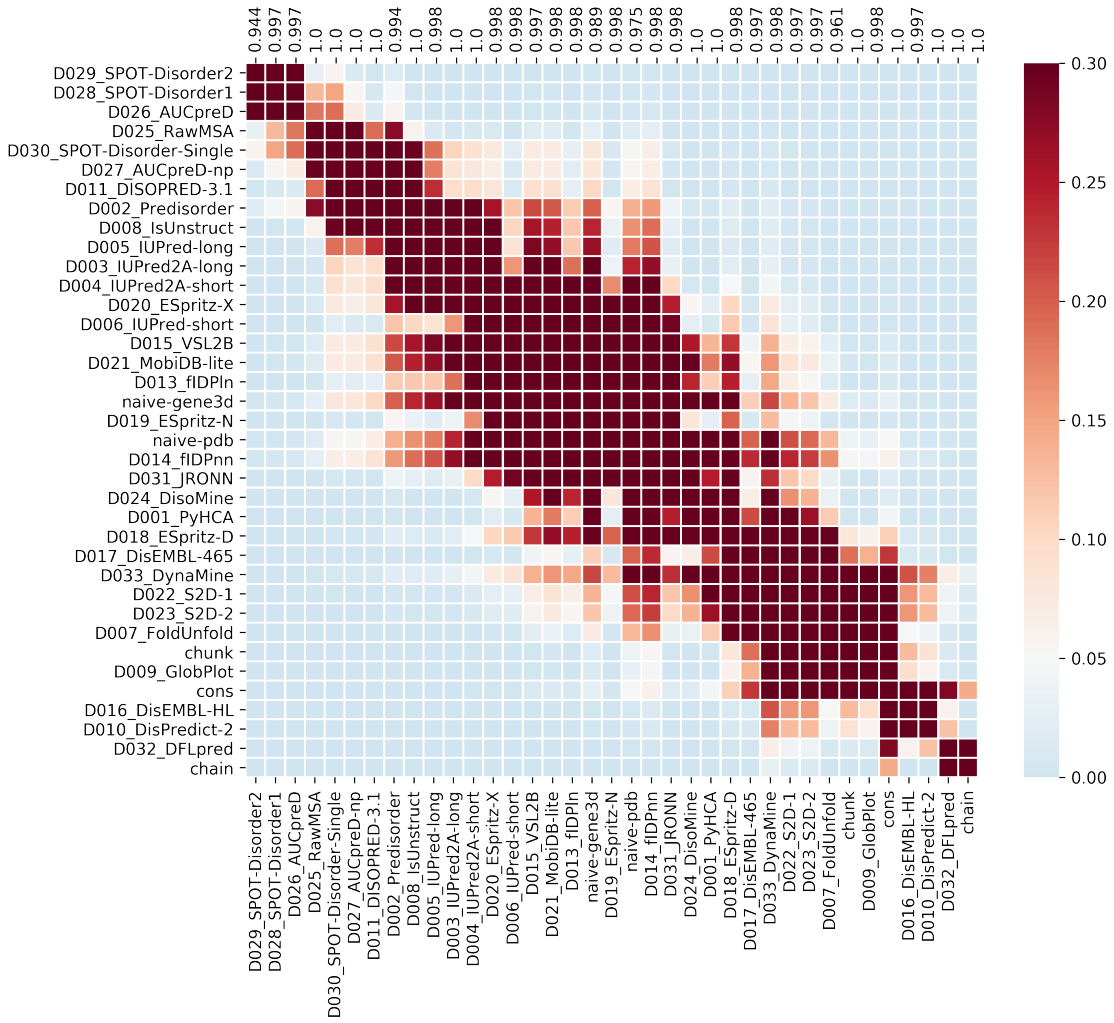
Plot of the AUCs from the ROC curve and PR curve



ROC AUC on the x axis, Precision-Recall (PR) AUC on the y axis. ROC AUC is calculated including ROC curve points with x values (FPR) between 0 and 0.2. PR AUC is calculated including PR curve points with x values (Recall) between 0.8 and 1. Both AUCs are then rescaled to the [0, 1] range.

1.2.9 Average overall ranking

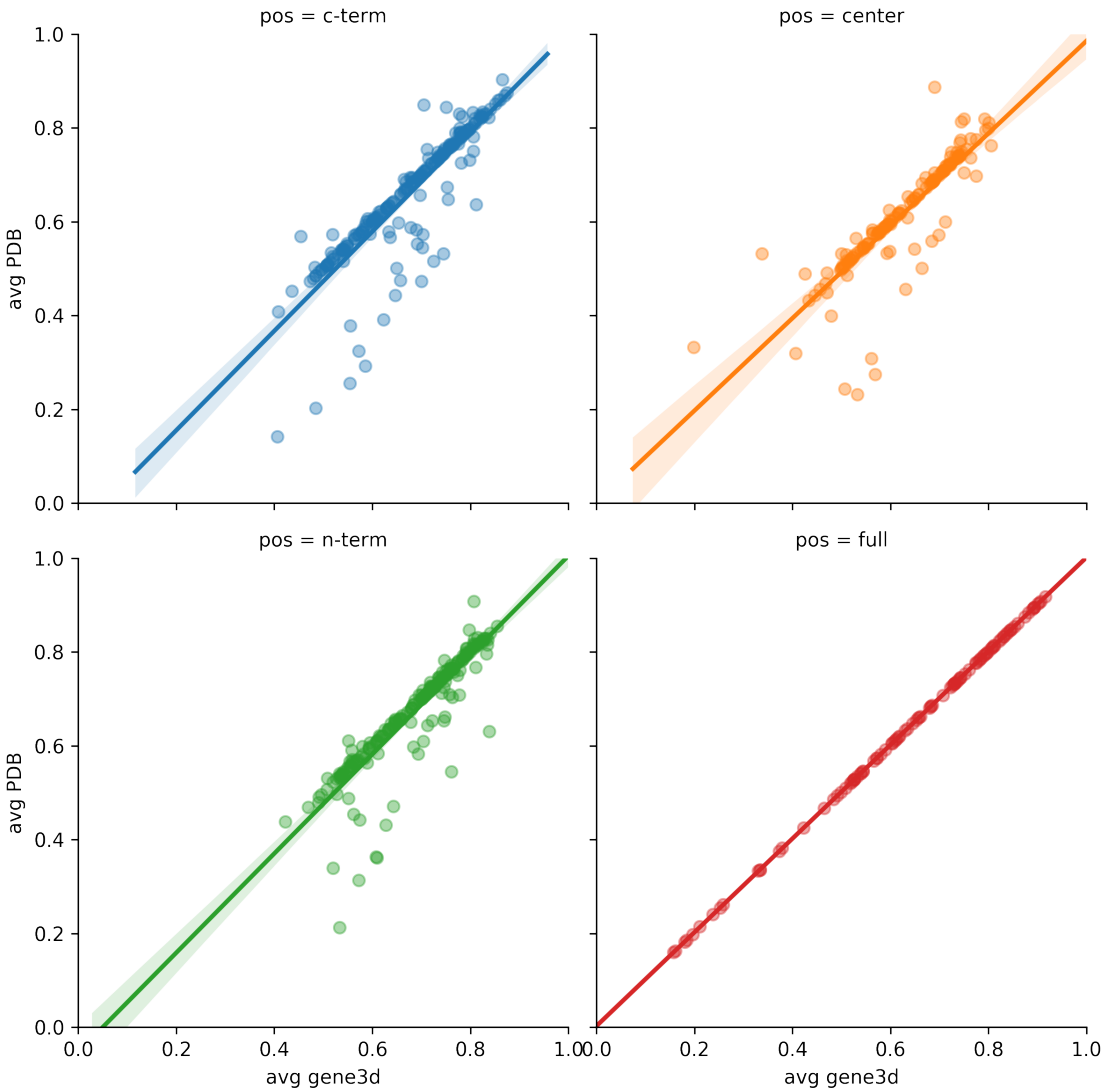
Predictor ranking and ranking statistical significance. Predictors are ranked on average rank from metrics scores: Balanced Accuracy, MCC, Precision, Recall, F1-score, F1-scores on negatives, Precision on negatives, Specificty, ROC AUC, PR AUC.



Heatmap of the p-value associated to the statistical significance of the difference between ranking distributions. Coloramap is centered on 0.05 so that any pvalue above 0.05 is **red**-ish. Red color indicates that the ranking difference between two predictors is not statistically significant. Top tick labels of *x* axis display prediction coverage for each predictor.

1.2.10 Accuracy correlation between datasets

Per target average balanced accuracy correlation between *simple* and *pdb* negative definition. Datasets is divided by average disorder position in targets.



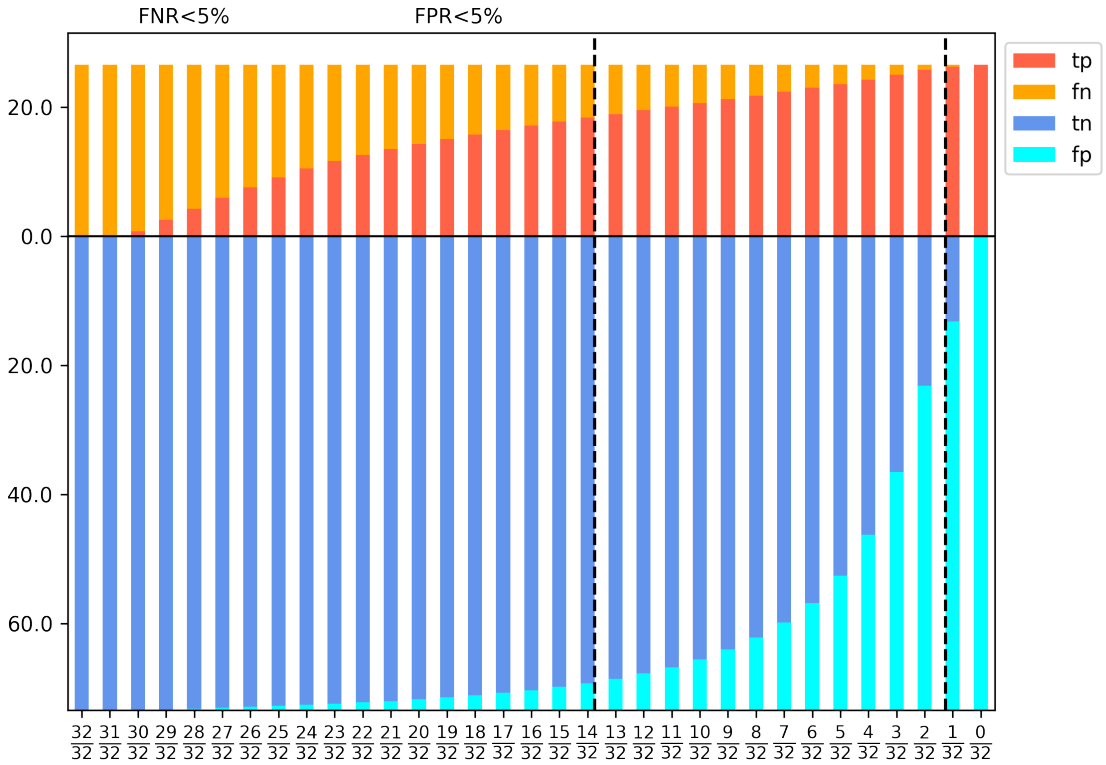
Average balanced accuracy for targets with reference negative defined by the *simple* rule on x axis. Average balanced accuracy for targets with reference negative defined by the *pdb* rule on y axis. Each panel includes only targets with a specific average disorder position (C-terminal, N-Terminal, central, full-disorder)

1.3 Consensus

Consensus among all prediction methods was calculated as the fraction of positive predictions per residue.

1.3.1 Confusion matrix per threshold

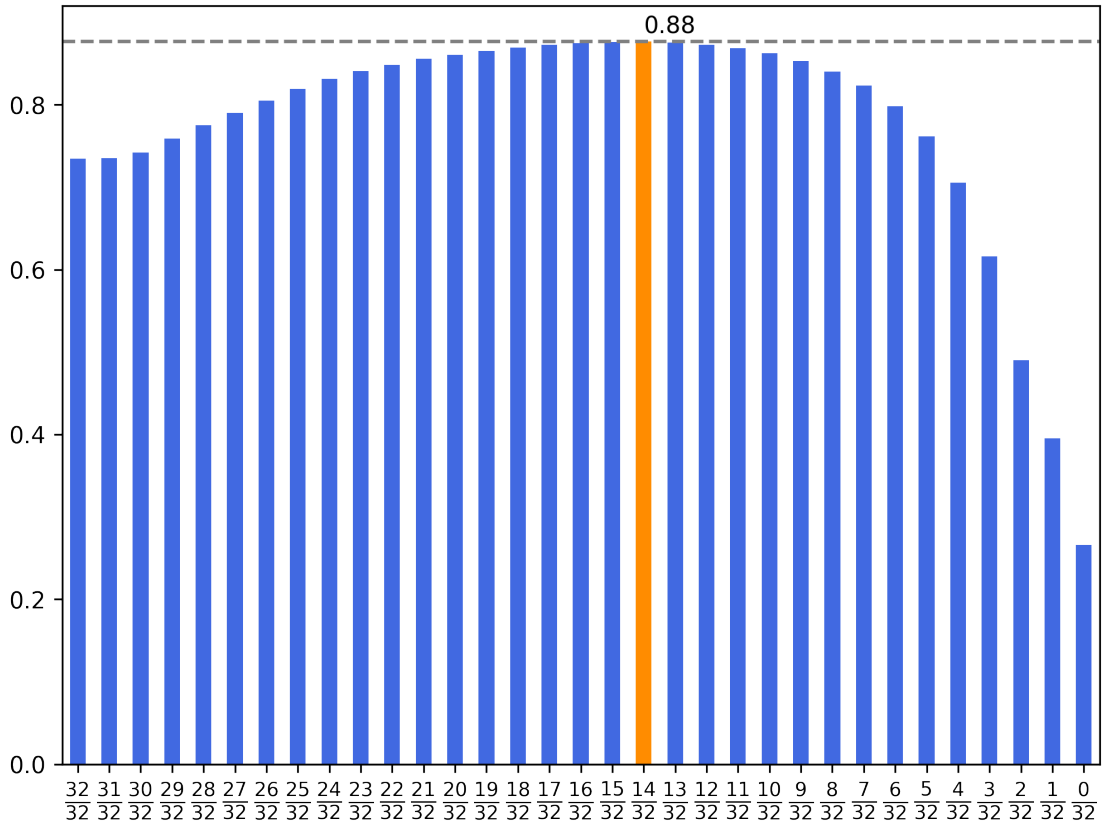
Predicted and actual positive and negatives for each threshold on the consensus score.



Percentage of correct and wrong assignment of positives (above 0) and negatives (below 0) for each threshold of the consensus.

1.3.2 Accuracy per threshold

Balanced accuracy score for each threshold of the consensus.



Accuracy distribution for each consensus threshold. Bar of max threshold is highlighted in orange.

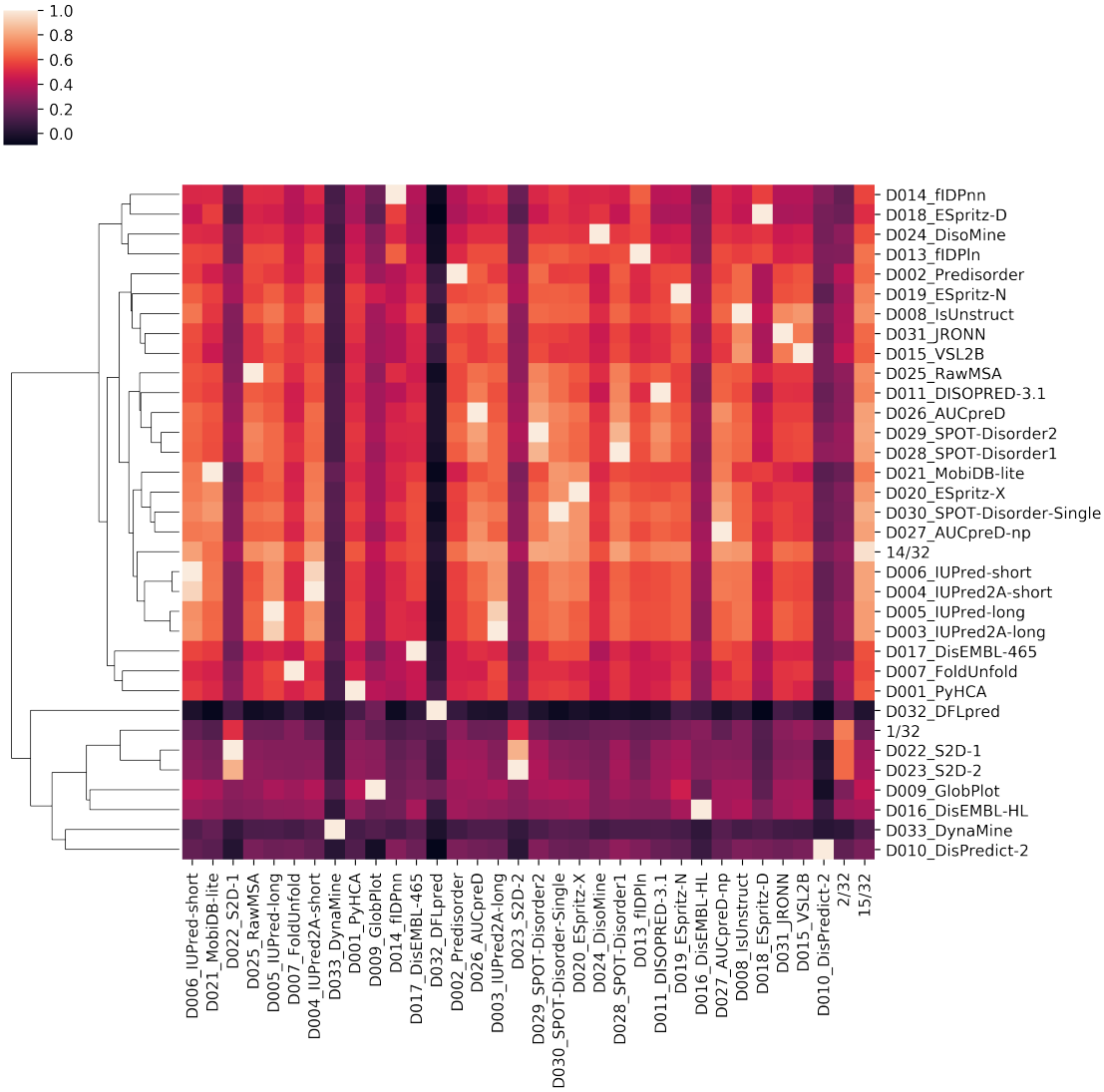
1.3.3 Percentage of correct/incorrect classifications

Percentage of correct and incorrect classifications for nositives (defined by DisProt), negatives (defined by PDB) and undefined residues for each predictor.

	DisProt		PDB		Undefined	
	TP	FN	TN	FP	TN	FP
D006_IUPred-short	56.2	43.8	94.7	5.3	66.2	33.8
D021_MobiDB-lite	41.1	58.9	98.6	1.4	77.2	22.8
D022_S2D-1	86.7	13.3	47.3	52.7	22.3	77.7
D025_RawMSA	68.2	31.8	93.1	6.9	60.6	39.4
D005_IUPred-long	64.5	35.5	92.5	7.5	58.9	41.1
D007_FoldUnfold	66.6	33.4	82.5	17.5	48.8	51.2
D004_IUPred2A-short	56.3	43.7	94.9	5.1	66.3	33.7
D033_DynaMine	1.7	98.3	100.0	0.0	97.9	2.1
D001_PyHCA	65.8	34.2	84.8	15.2	53.7	46.3
D009_GlobPlot	38.3	61.7	90.0	10.0	70.0	30.0
D014_fIDPnn	30.0	70.0	98.7	1.3	93.0	7.0
D017_DisEMBL-465	38.7	61.3	95.2	4.8	76.1	23.9
D032_DFLpred	9.4	90.6	89.0	11.0	89.4	10.6
D002_Predisorder	80.7	19.3	82.4	17.6	41.8	58.2
D026_AUCpreD	68.7	31.3	95.5	4.5	54.0	46.0
D003_IUPred2A-long	63.4	36.6	92.8	7.2	59.6	40.4
D023_S2D-2	85.8	14.2	50.4	49.6	24.7	75.3
D029_SPOT-Disorder2	75.9	24.1	95.1	4.9	45.8	54.2
D030_SPOT-Disorder-Single	55.6	44.4	97.7	2.3	65.3	34.7
D020_ESpritz-X	53.2	46.8	95.6	4.4	67.5	32.5
D024_DisoMine	57.4	42.6	91.0	9.0	67.3	32.7
D028_SPOT-Disorder1	74.8	25.2	94.2	5.8	48.8	51.2
D013_fIDPln	50.5	49.5	94.6	5.4	79.3	20.7
D011_DISOPRED-3.1	64.5	35.5	93.9	6.1	49.9	50.1
D019_ESpritz-N	68.7	31.3	86.6	13.4	51.6	48.4
D016_DisEMBL-HL	53.0	47.0	74.4	25.6	63.8	36.2
D027_AUCpreD-np	57.3	42.7	96.5	3.5	65.1	34.9
D008_IsUnstruct	74.8	25.2	86.0	14.0	49.3	50.7
D018_ESpritz-D	35.2	64.8	95.6	4.4	88.4	11.6
D031_JRONN	74.1	25.9	81.7	18.3	47.3	52.7
D015_VSL2B	81.5	18.5	77.2	22.8	40.0	60.0
D010_DisPredict-2	41.6	58.4	81.6	18.4	72.5	27.5

1.3.4 clustermap of binary predictions correlation

Correlation of binary states between predictors.



Heatmap of the correlation of binary prediction states for each couple of predictors. Pearson R is calculated between all predictions. Clustering is based on Euclidean distance calculated over an array (column) of R correlation coefficients.

1.4 Fully disordered targets

Statistics calculated for the subset of targets that are reported as completely disordered in DisProt.

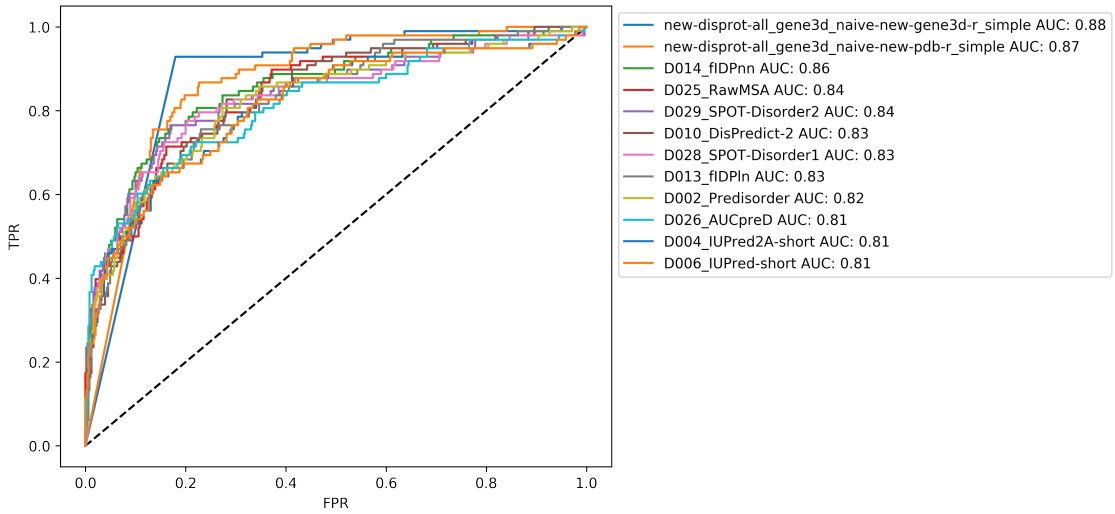
1.4.1 Correctly and incorrectly classified full IDPs

Number of correctly and incorrectly classified full IDPs with a prediction tolerance of 5%.

Actual	Positives		Negatives	
Predicted	TP	FN	FP	TN
random_chain	104	0	0	542
D001_PyHCA	104	0	1	541
random_chunk	104	0	2	540
D023_S2D-2	104	0	12	530
D015_VSL2B	104	0	20	522
D022_S2D-1	104	0	30	512
cons	104	0	120	422
D019_ESpritz-N	103	1	1	541
D002_Predisorder	102	2	3	539
D008_IsUnstruct	102	2	8	534
naive-gene3d	102	2	101	441
D004_IUPred2A-short	101	3	1	541
D026_AUCpreD	101	3	6	536
D006_IUPred-short	100	4	0	542
D009_GlobPlot	100	4	0	542
D027_AUCpreD-np	100	4	5	537
D020_ESpritz-X	99	5	3	539
D031_JRONN	99	5	5	537
D029_SPOT-Disorder2	99	5	11	531
D028_SPOT-Disorder1	99	5	16	526
naive-pdb	99	5	75	467
D016_DisEMBL-HL	98	6	2	540
D017_DisEMBL-465	97	7	0	542
D010_DisPredict-2	97	7	14	528
D024_DisoMine	96	8	44	498
D011_DISOPRED-3.1	95	9	3	539
D025_RawMSA	95	9	14	528
D003_IUPred2A-long	94	10	2	540
D005_IUPred-long	94	10	4	538
D007_FoldUnfold	94	10	127	415
D030_SPOT-Disorder-Single	93	11	3	539
D013_fIDPln	93	11	34	508
D014_fIDPnn	89	15	11	531
D021_MobiDB-lite	76	28	1	541
D018_ESpritz-D	75	29	45	497
D032_DFLpred	39	65	0	542
D033_DynaMine	28	76	0	542

1.4.2 Full IDPs ROC

ROC for the classification power of Full IDPs. Average disorder scores for each target is compared to full IDPs (positives) and partial IDPs (negatives). 5% prediction tolerance is applied.



FPR on the x axis, TPR on the y axis. Methods are sorted by their AUC. Only first 12 methods are shown.