

CAID

June 24, 2019

# Contents

<b>1</b>	<b>CAID</b>	<b>2</b>
1.1	Dataset	2
1.2	Evaluation metrics	2
1.2.1	Balanced accuracy	3
1.2.2	F1-score	4
1.2.3	MCC	4
1.2.4	Per target accuracy	5
1.2.5	Target correlation matrix	6
1.2.6	ROC curve	7
1.2.7	PR curve	8
1.2.8	pROC/pPR scatter plot	8
1.2.9	Average overall ranking	9
1.2.10	Accuracy correlation between datasets	10
1.3	Consensus	11
1.3.1	Confusion matrix per threshold	11
1.3.2	Accuracy per threshold	12
1.3.3	Percentage of correct/incorrect classifications	13
1.3.4	clustermap of binary predictions correlation	14
1.4	Fully disordered targets	15
1.4.1	Correctly and incorrectly classified full IDPs	15
1.4.2	Full IDPs ROC	16

# Chapter 1

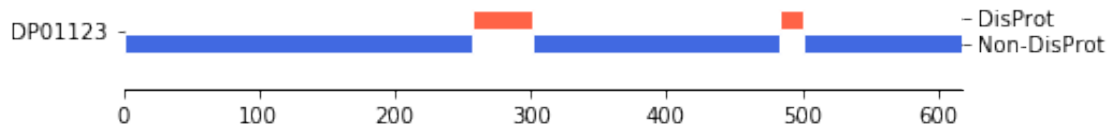
## CAID

### 1.1 Dataset

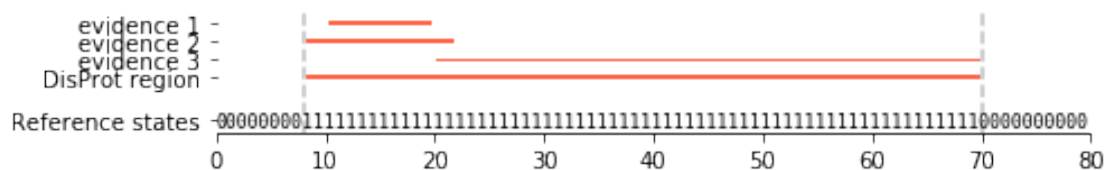
Critical Assessment of Intrinsic Disorder (CAID) is a continuous experiment where prediction methods for intrinsic disorder (ID) are blind tested on unpublished DisProt data.

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

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



### 1.2 Evaluation metrics

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

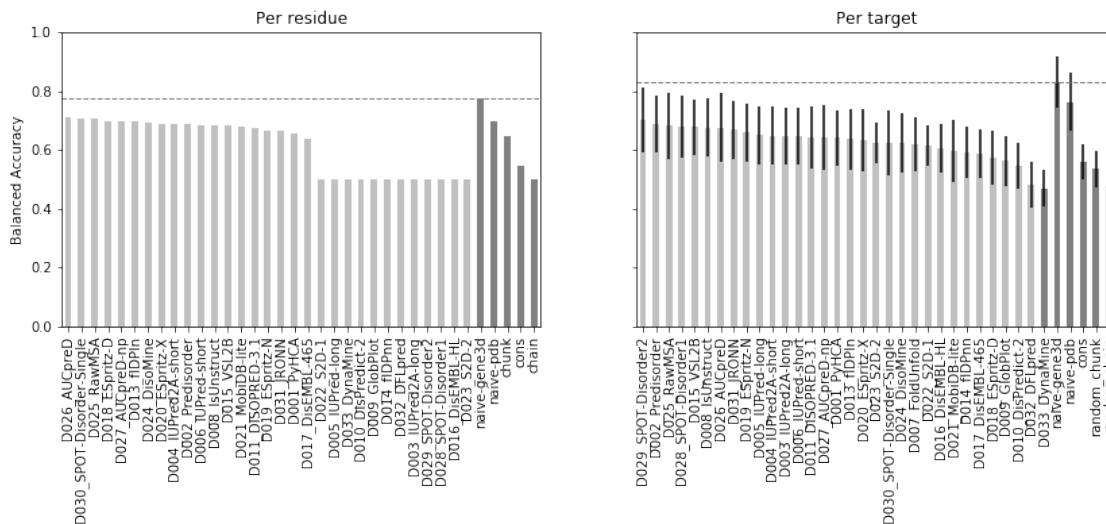
	FN	FP	TN	TP	BAC	F1s	MCC	Pre	Rec	Rec_n	AUC_PRC	AUC_ROC	npred	nref
naive-gene3d	0	126335	155656	54604	0.776	0.464	0.408	0.302	1.000	0.552	0.616	0.772	646	646
D026_AUCpreD	12478	96681	177823	42077	0.710	0.435	0.316	0.303	0.771	0.648	0.479	0.757	644	646
D030_SPOT-Disorder-Single	14078	92074	189917	40526	0.708	0.433	0.314	0.306	0.742	0.673	0.318	0.757	646	646
D025_RawMSA	18661	70521	211470	35943	0.704	0.446	0.324	0.338	0.658	0.750	0.414	0.780	646	646
naive-pdb	0	171438	110553	54604	0.696	0.389	0.308	0.242	1.000	0.392	0.588	0.688	646	646
D018_ESpritz-D	18155	78030	203889	36405	0.695	0.431	0.304	0.318	0.667	0.723	0.410	0.774	645	646
D027_AUCpreD-np	17119	83858	198133	37485	0.695	0.426	0.299	0.309	0.686	0.703	0.428	0.751	646	646
D013_fIDPIn	22805	54094	227897	31713	0.695	0.452	0.330	0.370	0.582	0.808	0.422	0.793	645	646
D024_DisoMine	17779	81922	200069	36825	0.692	0.425	0.296	0.310	0.674	0.709	0.388	0.765	646	646
D002_Predisorder	19429	71553	196647	35068	0.688	0.435	0.300	0.329	0.643	0.733	0.325	0.747	642	646
D020_ESpritz-X	17422	86006	195913	37138	0.688	0.418	0.287	0.302	0.681	0.695	0.304	0.740	645	646
D004_IUPred2A-short	17789	83861	198130	36815	0.688	0.420	0.290	0.305	0.674	0.703	0.313	0.741	646	646
D006_IUPred-short	18852	80184	201735	35708	0.685	0.419	0.287	0.308	0.654	0.716	0.311	0.739	645	646
D008_IsUnstruct	19234	79292	202699	35370	0.683	0.418	0.285	0.308	0.648	0.719	0.323	0.744	646	646
D015_VSL2B	17253	89996	191453	37040	0.681	0.409	0.275	0.292	0.682	0.680	0.301	0.732	644	646
D021_MobiDB-lite	22254	66649	215270	32306	0.678	0.421	0.288	0.326	0.592	0.764	0.366	0.737	645	646
D011_DISOPRED-3.1	15249	106219	175772	39355	0.672	0.393	0.256	0.270	0.721	0.623	0.290	0.701	646	646
D031_JRONN	21714	76096	205823	32846	0.666	0.402	0.262	0.301	0.602	0.730	0.302	0.724	645	646
D019_ESpritz-N	20944	80018	201901	33616	0.666	0.400	0.259	0.296	0.616	0.716	0.296	0.714	645	646
D001_PyHCA	20332	88792	193199	34272	0.656	0.386	0.239	0.278	0.628	0.685	0.277	0.706	646	646
chunk	32511	32511	249479	22092	0.645	0.405	0.289	0.405	0.405	0.885	NaN	NaN	100	100
D017_DisEMBL-465	23573	82440	199009	30720	0.636	0.367	0.213	0.271	0.566	0.707	0.283	0.685	644	646
cons	12165	193161	88830	42439	0.546	0.292	0.074	0.180	0.777	0.315	0.193	0.567	646	646
D003_IUPred2A-long	54604	75	281916	0	0.500	0.000	-0.007	0.000	0.000	1.000	0.298	0.735	646	646
D022_S2D-1	54293	1	281448	0	0.500	0.000	-0.001	0.000	0.000	1.000	0.253	0.672	644	646
chain	45733	45733	236257	8870	0.500	0.162	0.000	0.162	0.162	0.838	NaN	NaN	100	100
D010_DisPredict-2	54604	1	281990	0	0.500	0.000	-0.001	0.000	0.000	1.000	0.250	0.637	646	646
D014_fIDPnn	54518	4	281987	0	0.500	0.000	-0.002	0.000	0.000	1.000	0.475	0.814	645	646
D005_IUPred-long	54560	85	281834	0	0.500	0.000	-0.007	0.000	0.000	1.000	0.298	0.737	645	646
D028_SPOT-Disorder1	54293	2	281447	0	0.500	0.000	-0.001	0.000	0.000	1.000	0.268	0.744	644	646
D023_S2D-2	54293	1	281448	0	0.500	0.000	-0.001	0.000	0.000	1.000	0.229	0.654	644	646
D029_SPOT-Disorder2	47751	4	221404	0	0.500	0.000	-0.002	0.000	0.000	1.000	0.340	0.760	610	646
D016_DisEMBL-HL	54293	1	281448	0	0.500	0.000	-0.001	0.000	0.000	1.000	0.274	0.654	644	646
D009_GlobPlot	54560	2	281917	0	0.500	0.000	-0.001	0.000	0.000	1.000	0.231	0.624	645	646
D033_DynaMine	54560	1	281918	0	0.500	0.000	-0.001	0.000	0.000	1.000	0.271	0.707	645	646
D032_DFLpred	54604	1	281990	0	0.500	0.000	-0.001	0.000	0.000	1.000	0.142	0.410	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

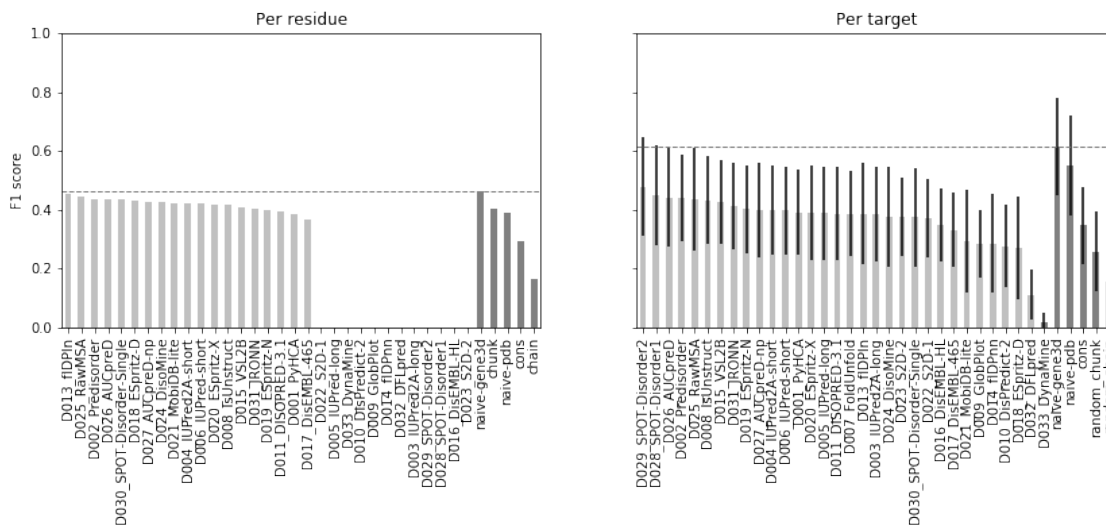
Comparison of predictors and baselines performance by balanced accuracy.



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

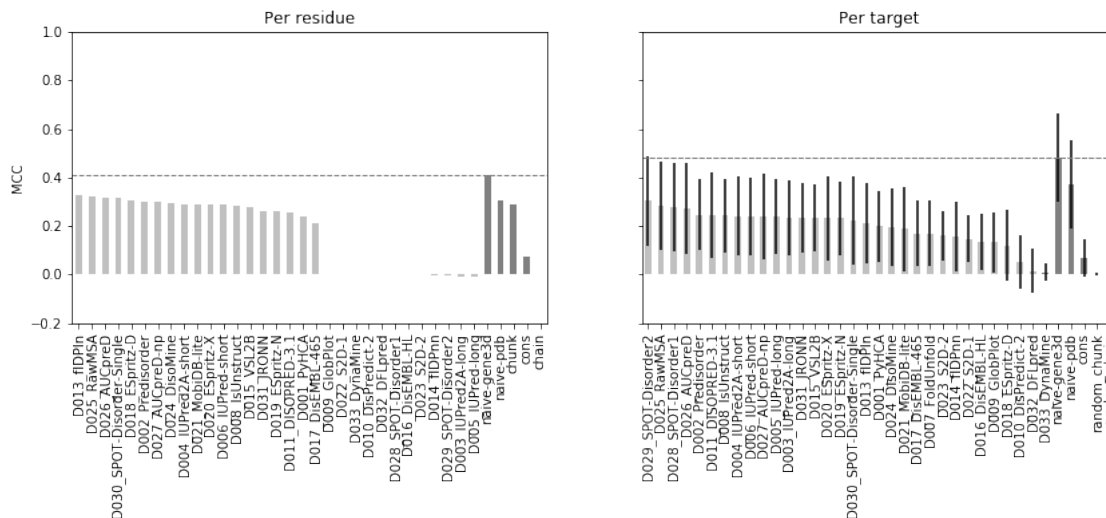
Comparison of predictors and baselines performance by F1-score.



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

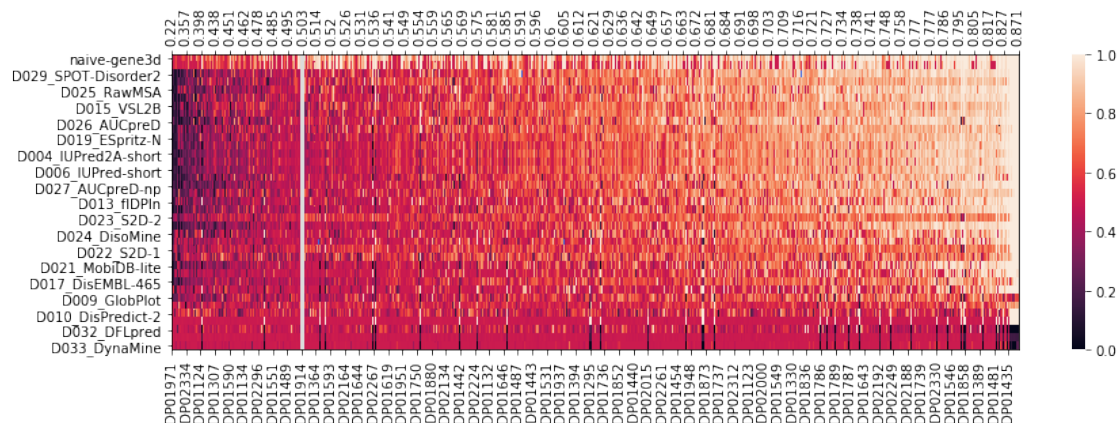
Comparison of predictors and baselines performance by Matthew's Correlation Coefficient.



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 baselines)



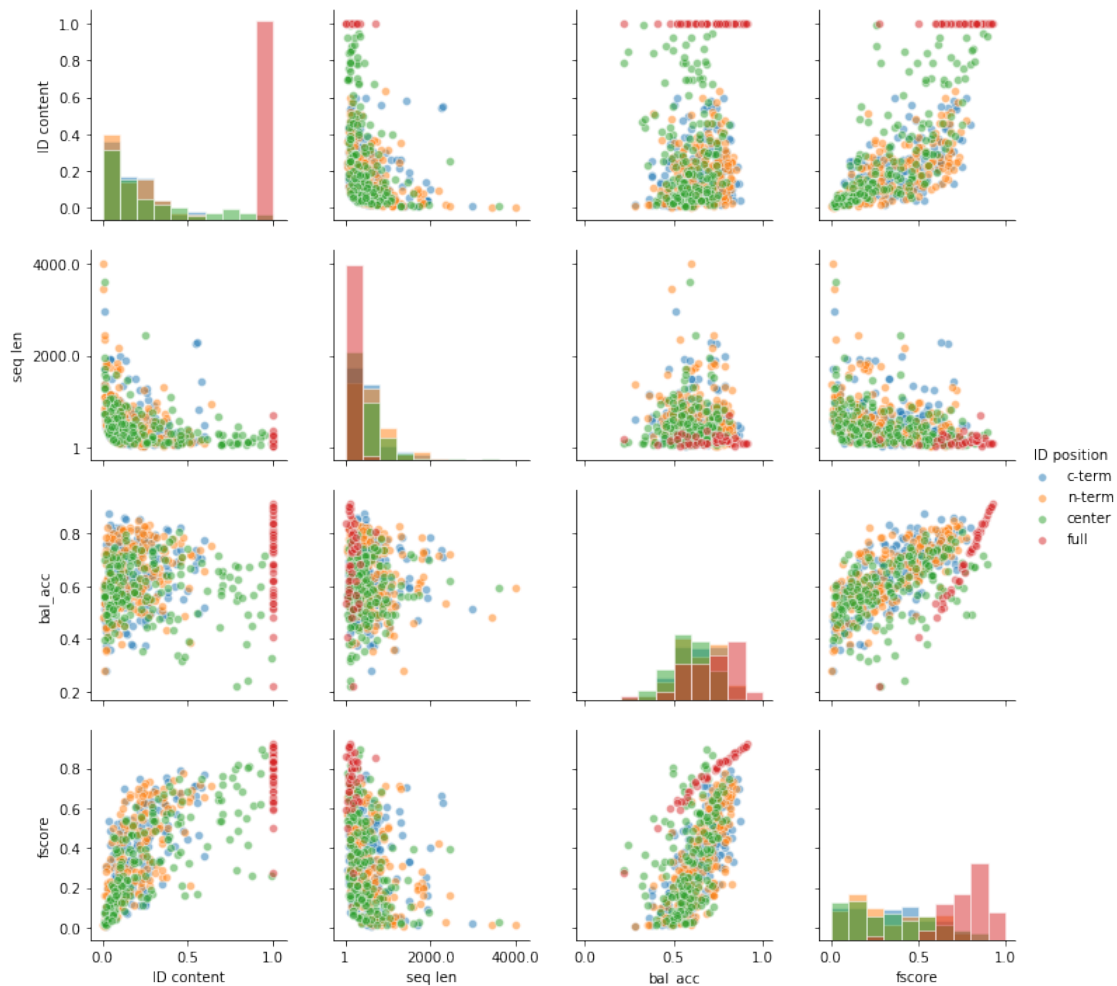
Heatmap of the predictors accuracy for each target. The higher the accuracy the lighter the color. Non-predicted targets are shown in blue.  $x$  and  $y$  axes are sorted by average accuracy over rows and columns respectively. A white semi-transparent vertical line marks the point where the average 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:

DP01971	DP01870	DP02010	DP01278	DP01898	DP01281
DP01432	DP01456	DP01128	DP01427	DP01130	DP02334
DP01494	DP01877	DP01203	DP01501	DP01584	DP01145

DP02149	DP01248	DP01498	DP01462	DP01124	DP01366
DP01512	DP01844	DP01139	DP01141	DP01407	DP01724
DP01907	DP01556	DP01316	DP01307	DP01925	DP01883
DP01185	DP01474	DP01878	DP02231	DP01163	DP01794
DP01477	DP02324	DP01590	DP01285	DP01774	DP02234
DP02328	DP01280	DP01500	DP01612	DP01978	DP02073
DP01999	DP01134	DP01647	DP01772	DP01434	DP01313
DP01324	DP01504	DP01172	DP01193	DP01428	DP02086
DP02296	DP01355	DP01503	DP01749	DP01499	DP01140
DP01323	DP01600	DP01869	DP01505	DP01887	DP01551
DP01150	DP01279	DP01473	DP02247	DP01854	DP02117
DP01152	DP02168	DP01170	DP01195	DP01489	DP01309
DP01196	DP02170	DP01177	DP01396	DP01252	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 if 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.

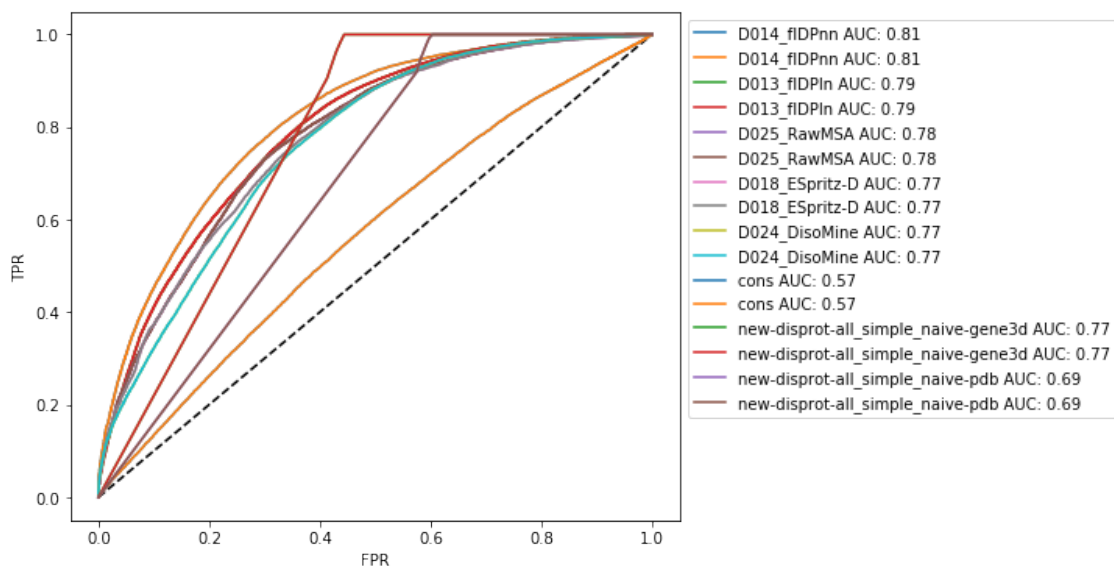


Correlation matrix of Balanced accuracy, F1-Score, Sequence length and ID content. Average position of disorder is color-coded. Figure matrix is symmetrical. Plots along diagonal axis display single variables distributions. No meaningful correlation is observed.

### 1.2.6 ROC curve

Receiver Operator Characteristic plot for predictors and baselines

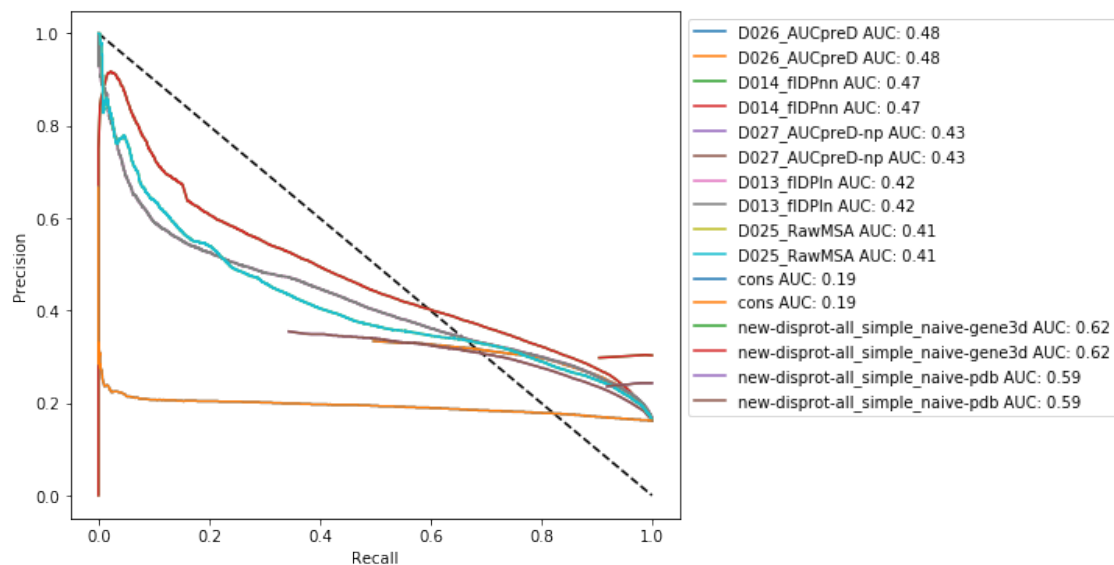




False Positive Rate (FPR) on  $x$  axis, True Positive Rate (TPR) on  $y$  axis. Methods are sorted by their Area Under the Curve (AUC). Only first ten methods plus baselines are shown.

### 1.2.7 PR curve

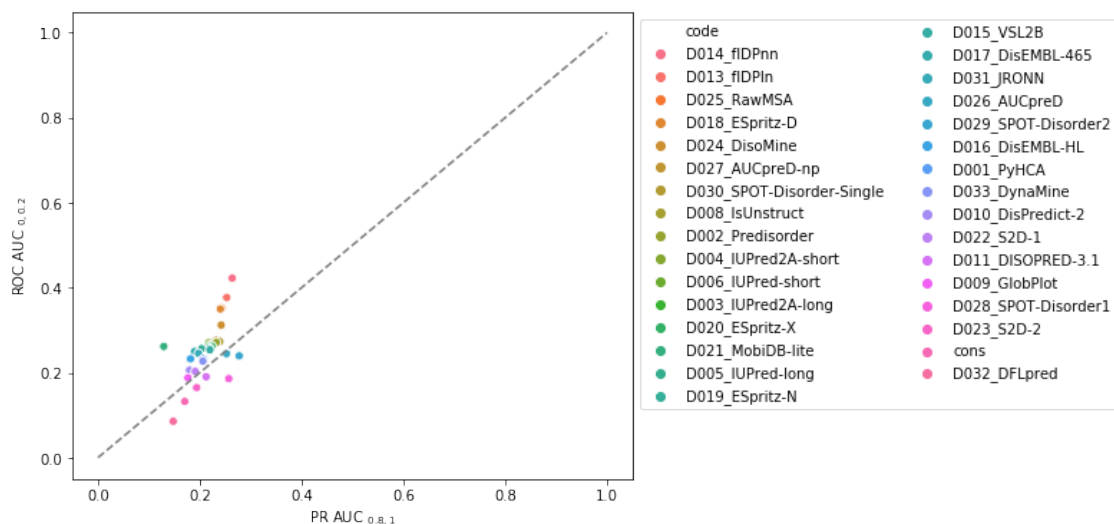
Precision Recall curve plot for predictors and baselines



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

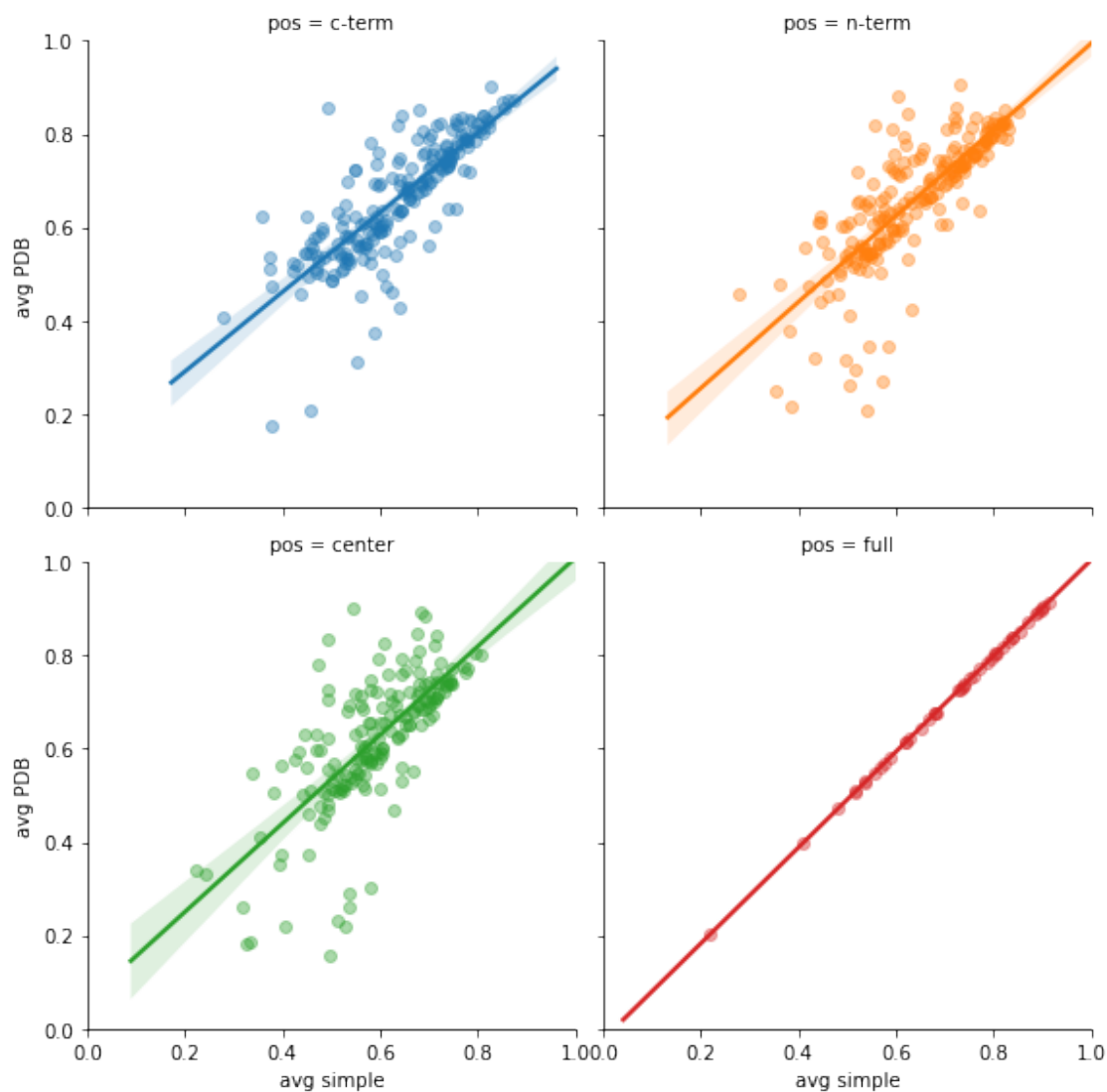


ROC AUC on the  $x$  axis, Precision-Recall (PR) AUC on the  $y$  axis. ROC AUC is calculated including ROC curve points with  $x$  values (FPR) between 0 and 0.2. PR AUC is calculated including PR curve points with  $x$  values (Recall) between 0.8 and 1. Both AUCs are then rescaled to the  $[0, 1]$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





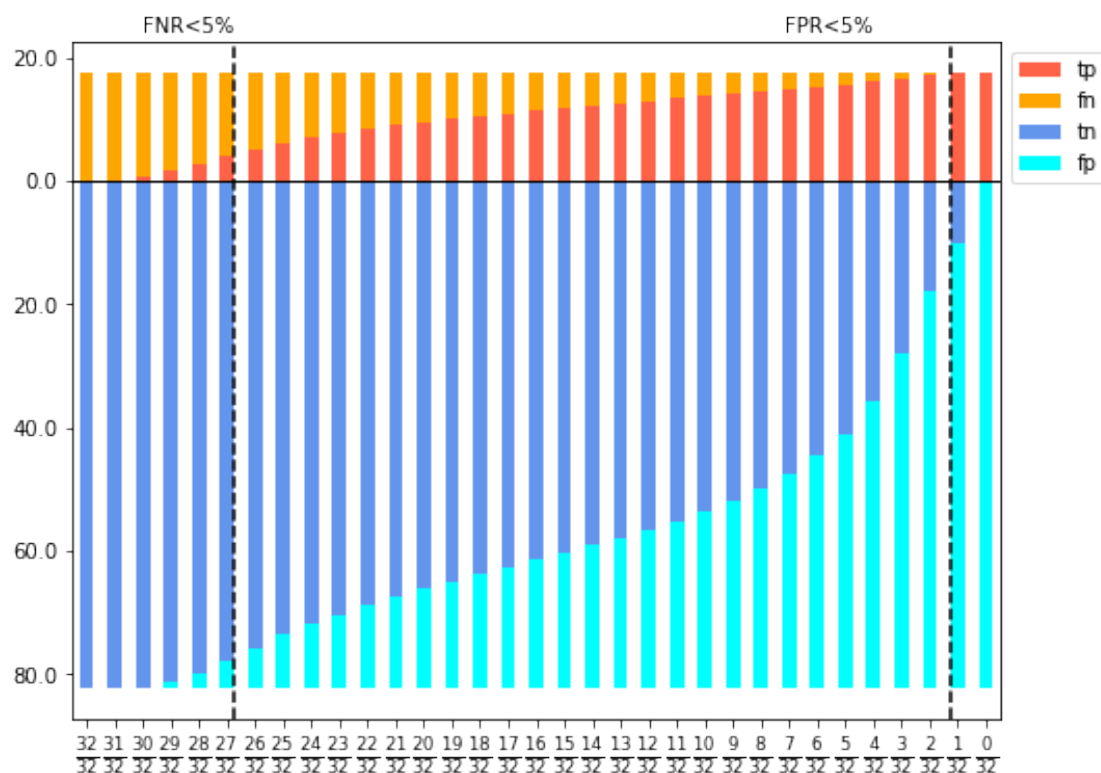
Average balanced accuracy for targets with reference negative defined by the *simple* rule on  $x$  axis. Average balanced accuracy for targets with reference negative defined by the *pdb* rule on  $y$  axis. Each panel includes only targets with a specific average disorder position (C-terminal, N-Terminal,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

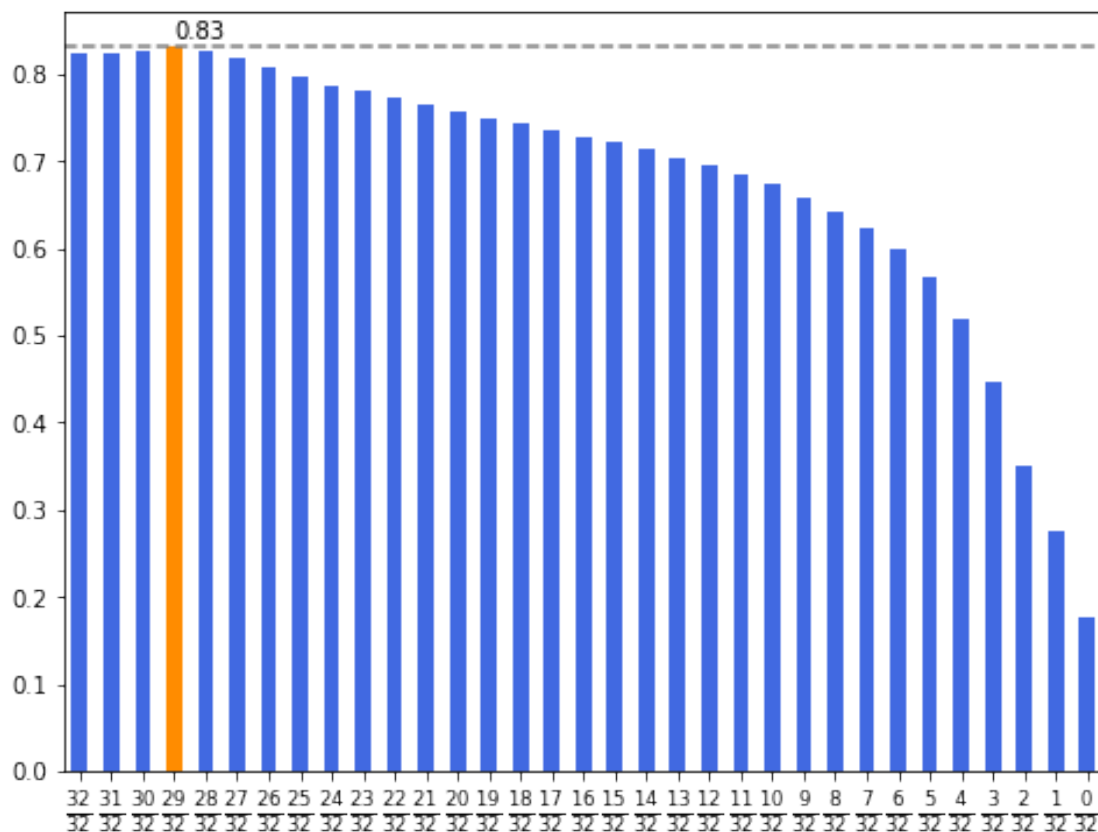
Predicted and actual positive and negatives for each threshold on the consensus score.



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.

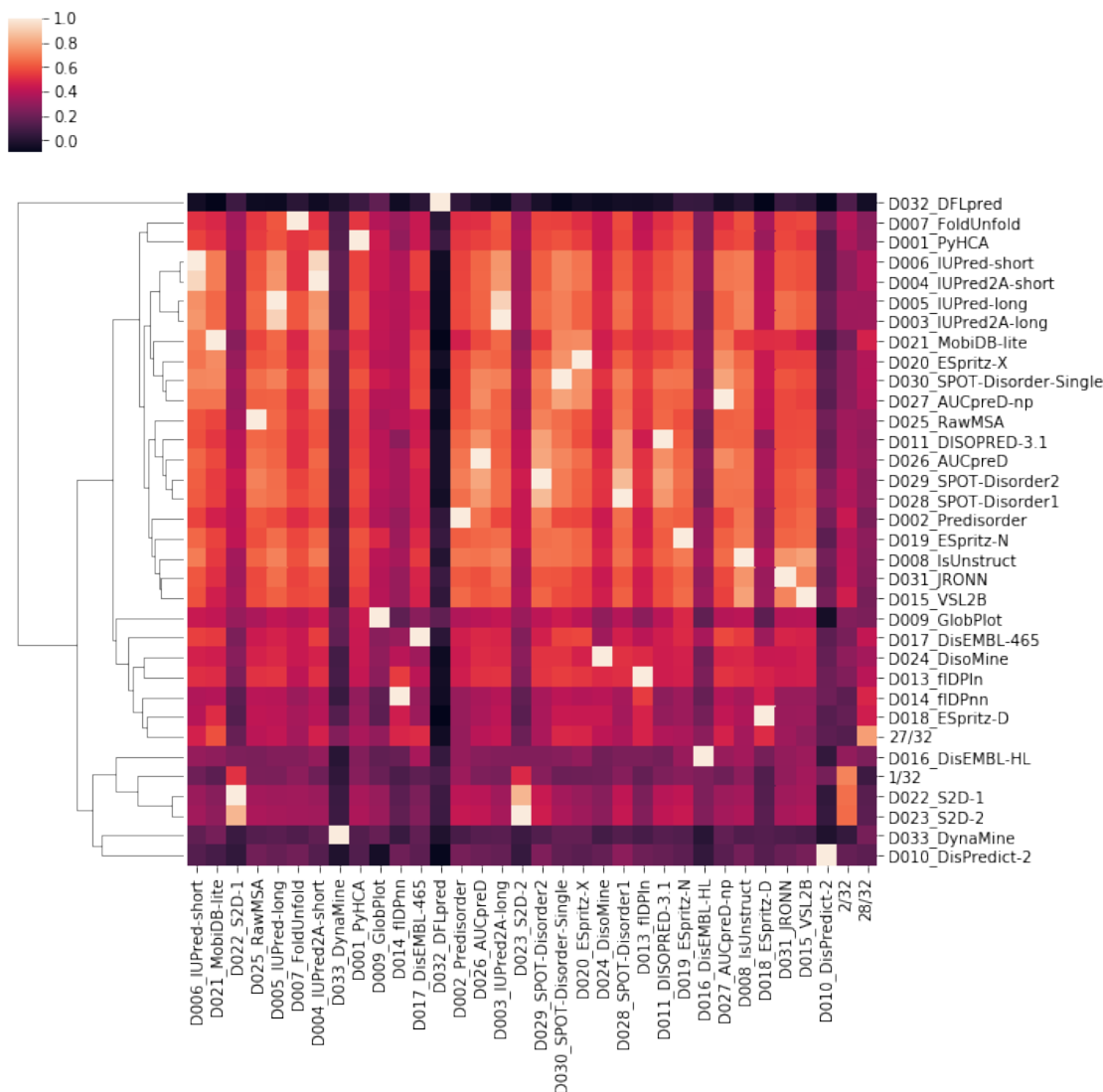
### 1.3.3 Percentage of correct/incorrect classifications

Percentage of correct and incorrect classifications for positives (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_fDPnn	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_fDPIn	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

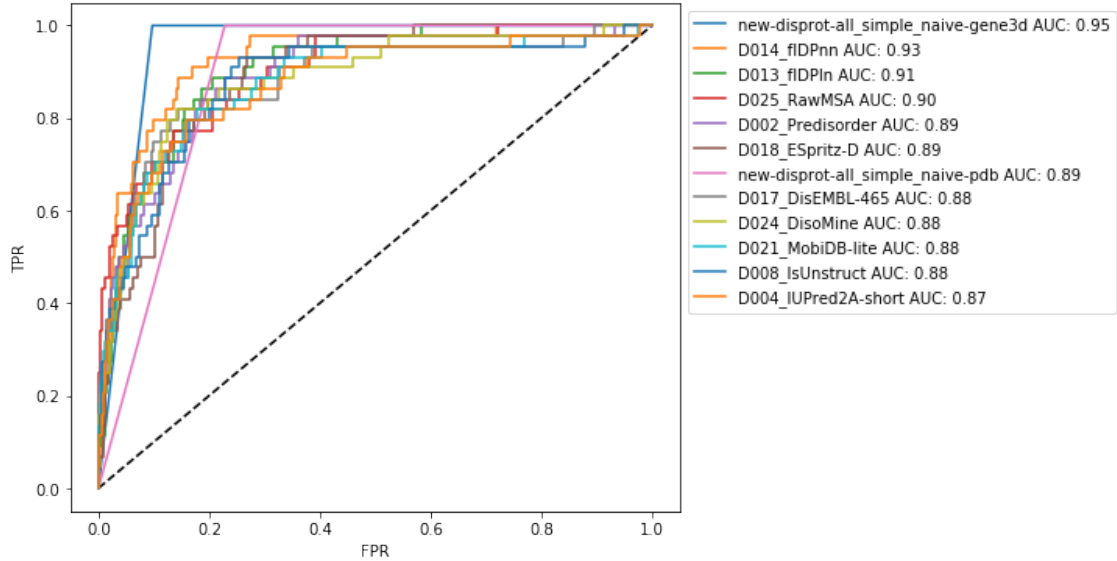
Number of correctly and incorrectly classified full IDPs with a prediction tolerance of 5%.



Actual Predicted	Positives		Negatives	
	TP	FN	FP	TN
D009_GlobPlot	44	0	0	602
D016_DisEMBL-HL	44	0	0	602
D017_DisEMBL-465	44	0	0	602
random_chain	44	0	0	602
D006_IUPred-short	44	0	1	601
random_chunk	44	0	1	601
D004_IUPred2A-short	44	0	2	600
D019_ESpritz-N	44	0	4	598
D001_PyHCA	44	0	5	597
D011_DISOPRED-3.1	44	0	6	596
D020_ESpritz-X	44	0	6	596
D030_SPOT-Disorder-Single	44	0	7	595
D027_AUCpreD-np	44	0	11	591
D002_Predisorder	44	0	13	589
D008_IsUnstruct	44	0	14	588
D026_AUCpreD	44	0	15	587
D015_VSL2B	44	0	24	578
D029_SPOT-Disorder2	44	0	28	574
D028_SPOT-Disorder1	44	0	31	571
D023_S2D-2	44	0	33	569
D024_DisoMine	44	0	55	547
D022_S2D-1	44	0	57	545
naive-gene3d	44	0	69	533
cons	44	0	160	442
naive-pdb	44	0	172	430
D010_DisPredict-2	43	1	18	584
D025_RawMSA	43	1	20	582
D013_fIDPIn	43	1	41	561
D003_IUPred2A-long	42	2	6	596
D005_IUPred-long	42	2	6	596
D031_JRONN	42	2	6	596
D014_fIDPnn	42	2	16	586
D018_ESpritz-D	42	2	52	550
D007_FoldUnfold	41	3	156	446
D021_MobiDB-lite	37	7	2	600
D032_DFLpred	13	31	0	602
D033_DynaMine	13	31	0	602

## 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.