

RNA-/DNA- binding sites prediction on protein

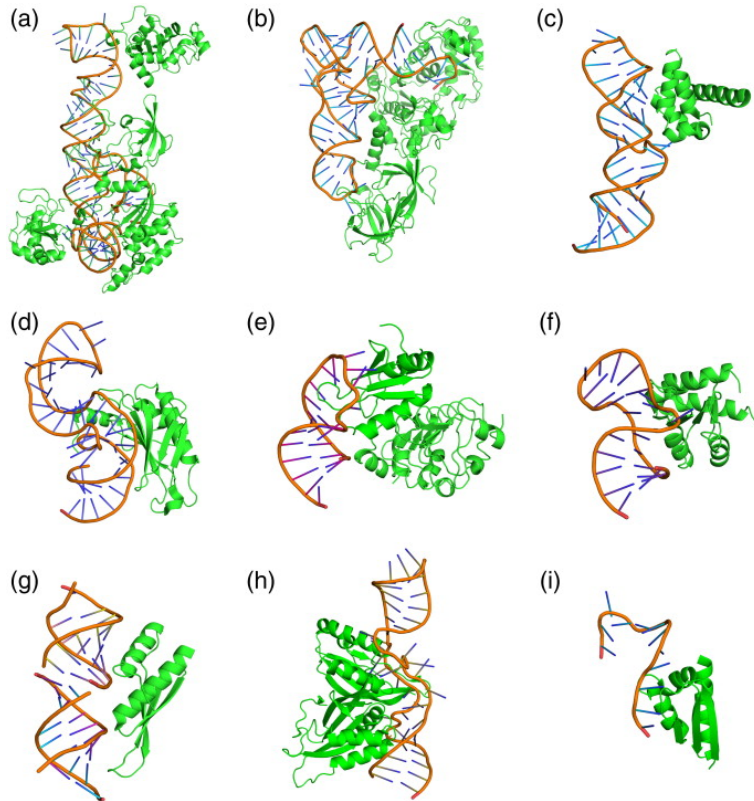
Zhichao (Chichau) Miao

Brazma lab – EBML-EBI

Teichmann lab – Wellcome Sanger Institute

Benasque 2018

Protein- Nucleic acid interactions



- Levels:
 - Molecular:
 - interaction
 - Residue:
 - binding sites
 - Structure:
 - complex structure
- Questions:
 - Driving force?
 - Features?
 - DNA-/RNA-binding difference?

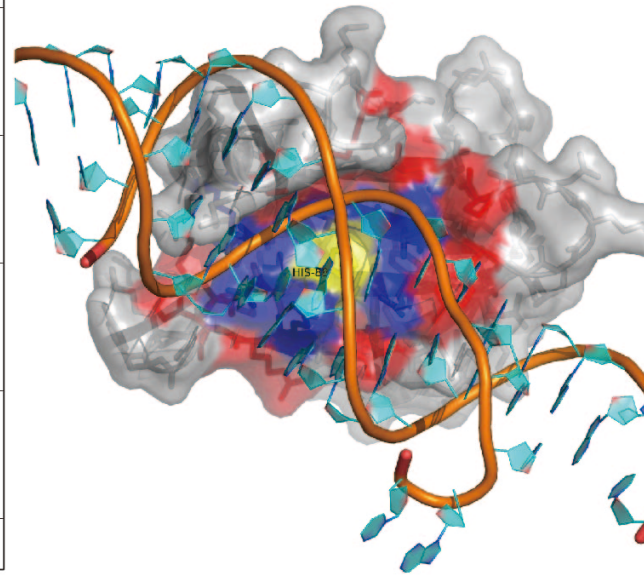
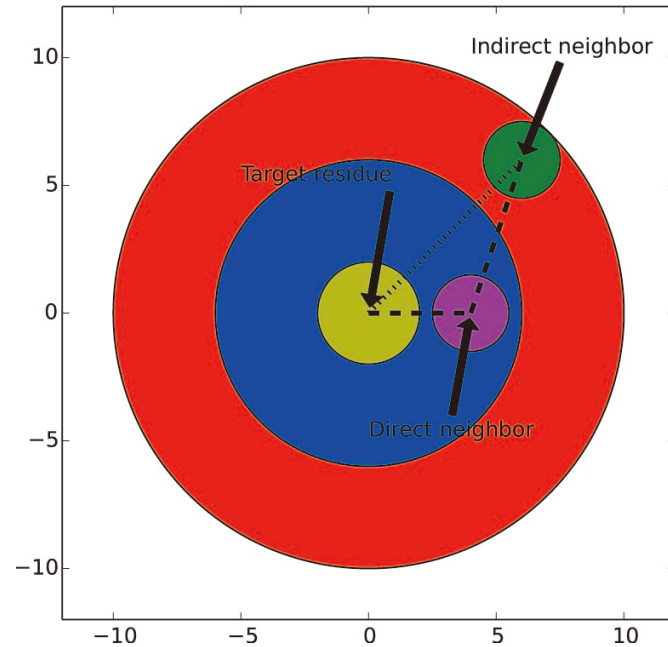
RBscore: binding sites prediction

- Based on 3D structure
- Simple features used: accessibility (ASA), electrostatics potential and sequence conservation entropy
- Surface grid based **neighboring network** integrates features in a linear approach.

$$E_{\text{pred}} = E_{\text{feat}} + \frac{\sum_{\text{direct neighbors}} u_{aa} \times f_{\text{neighbor type}} \times E_{\text{feat}}}{N_{\text{direct neighbors}}} + \frac{\sum_{\text{indirect neighbors}} v_{aa} \times g_{\text{neighbor type}} \times E_{\text{feat}}}{N_{\text{indirect neighbors}}}$$

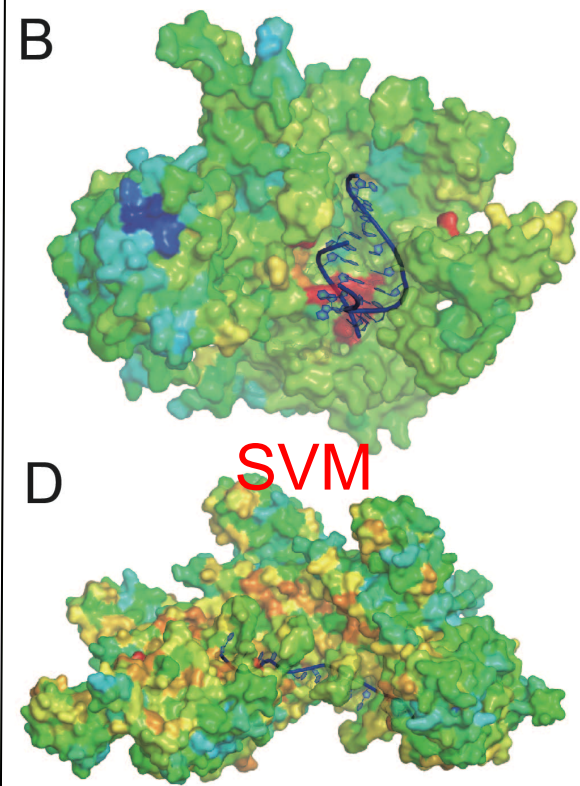
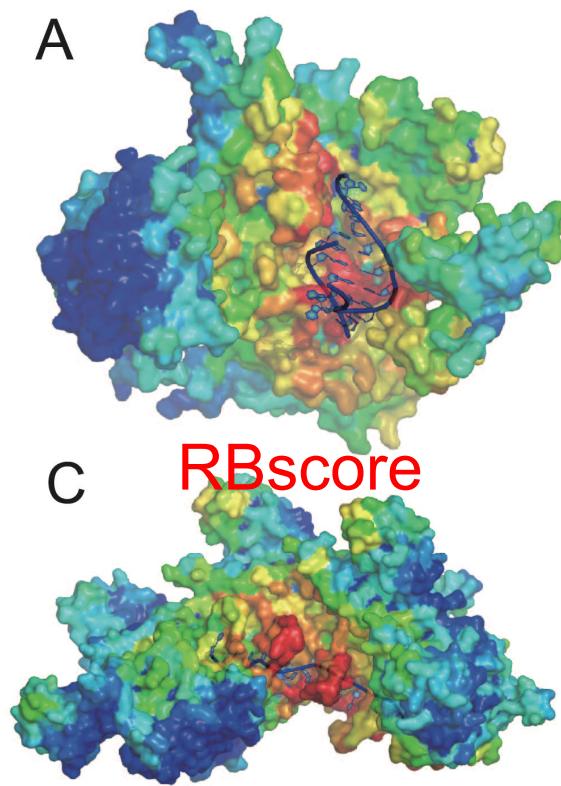
$$E_{\text{feat}} = w_{aa} \times \text{ASA} + \sum_{10} w_i \times \text{count}_{\text{ELEC}}(i) + w_{\text{CE}} \times \text{CE} + w_{\text{width}} \times \text{width} + C_{aa}$$

Why neighboring network?



- The binding interface is continuous.
- The RNA/DNA may slide into the binding pocket.

Energy funnel
plotted by
RBscore



State-of-the-art predictors

- Well-established problem...
- >5 papers, many without rigorous benchmarking
- >20 web server, programs (some do not work)
- >40 data sets

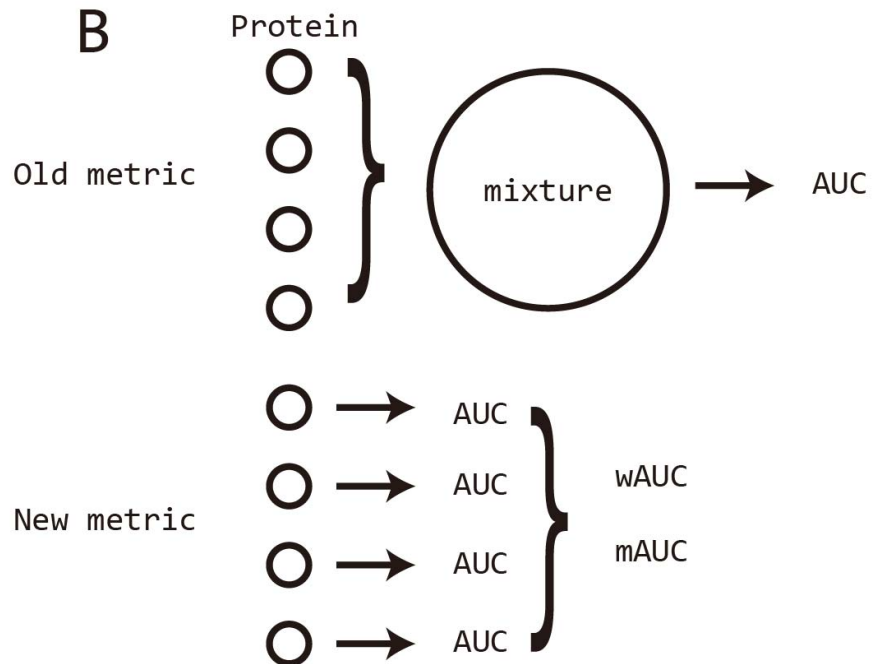
- DNA binding *vs.* RNA binding predictors
- Sequence based *vs.* structure based predictors

Biases in assessment

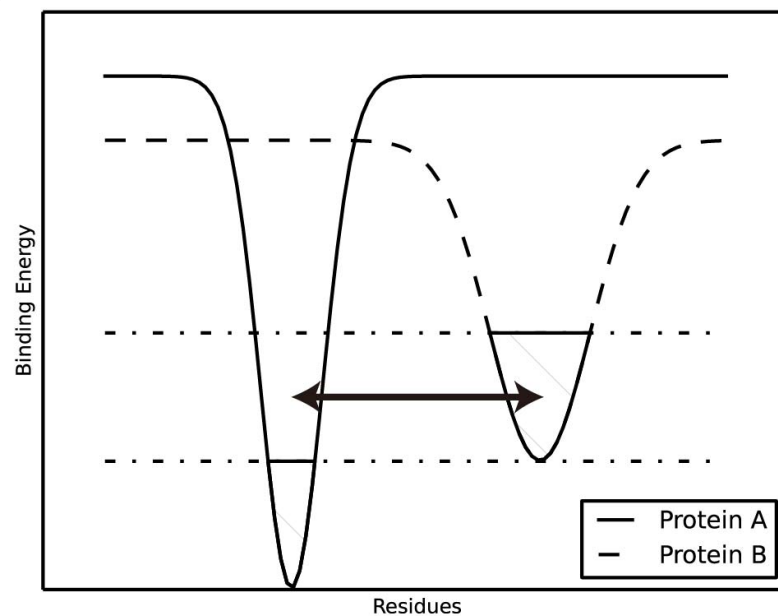
A

RPYACPVESCRRRFSRSDDELTRHIRIHTGQKPFQCRICMRNFSRSDHLTTHIRHTHTGEKPFACDICGRKFFARSDERKRHTKIHLR	#	cutoff(Å)
+	20	3.5
+	26	4.0
+	31	4.5
+	37	5.0
+	39	5.5
+	40	6.0

B

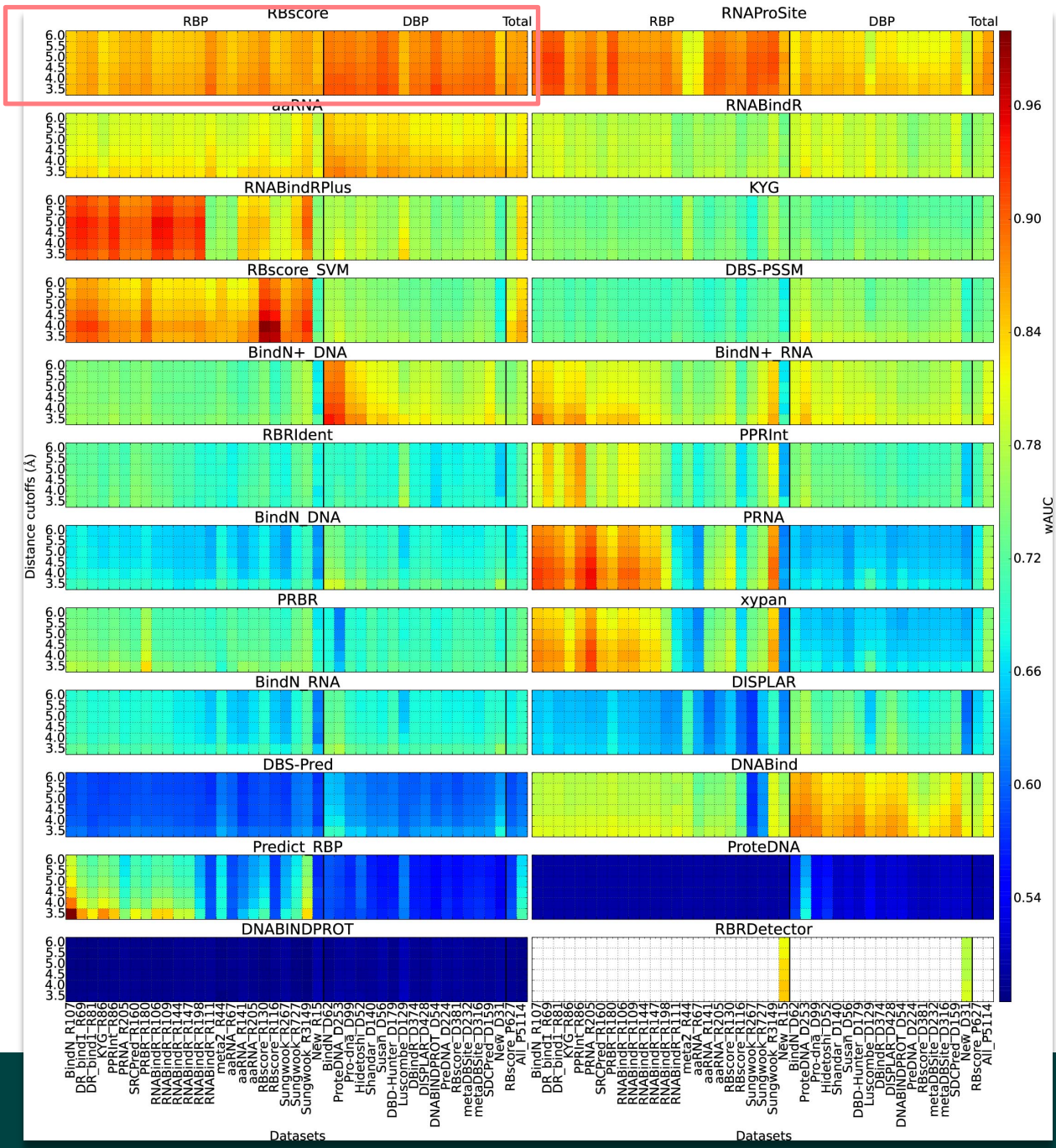


C



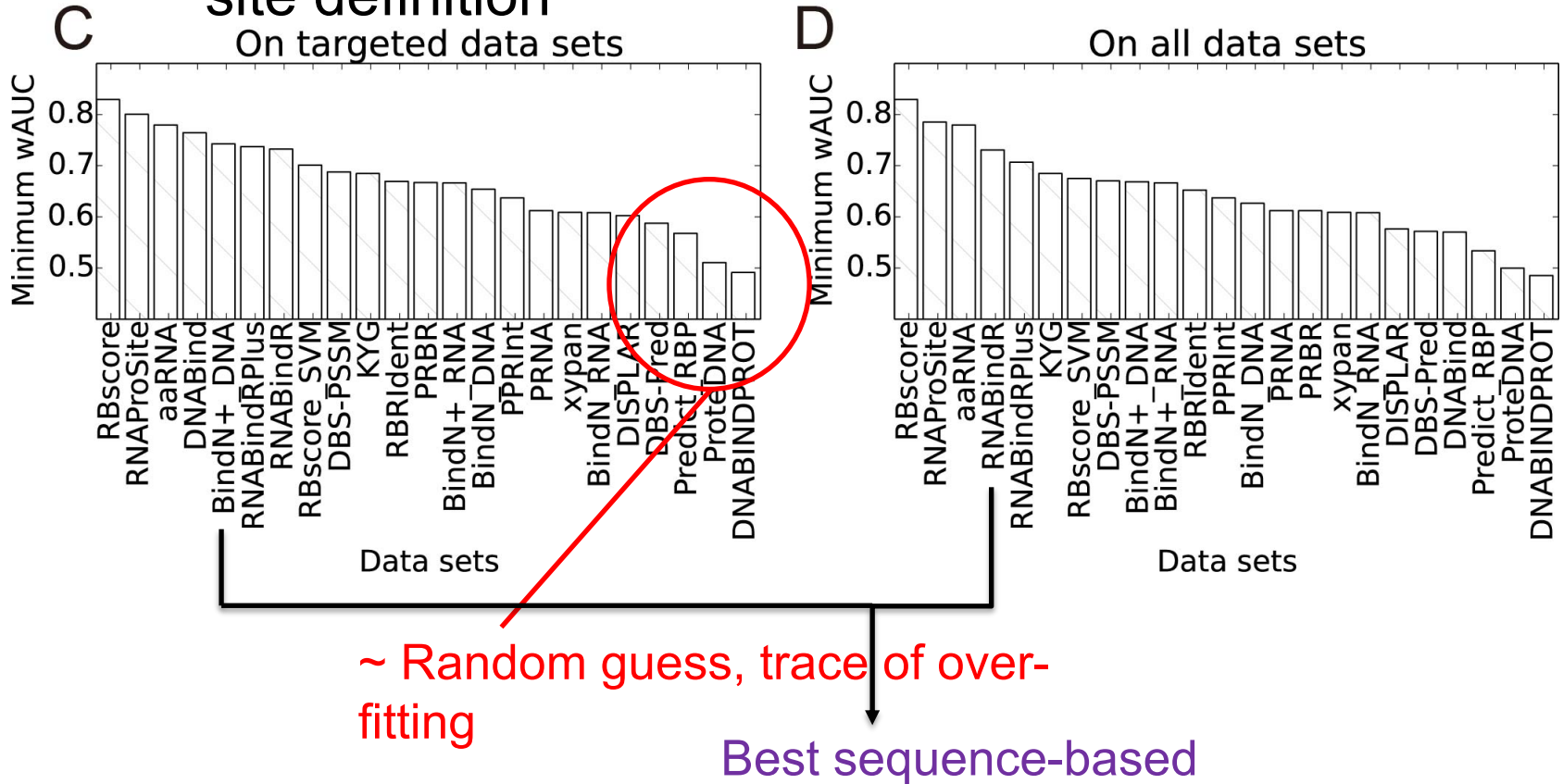
Benchmark summary

- 19 web servers and 3 stand-alone programs (25 approaches in total)
- 41 data sets
- Various metrics: wAUC, mAUC, binary criteria (AUC, PPV, SPC, SEN, F1, MCC)



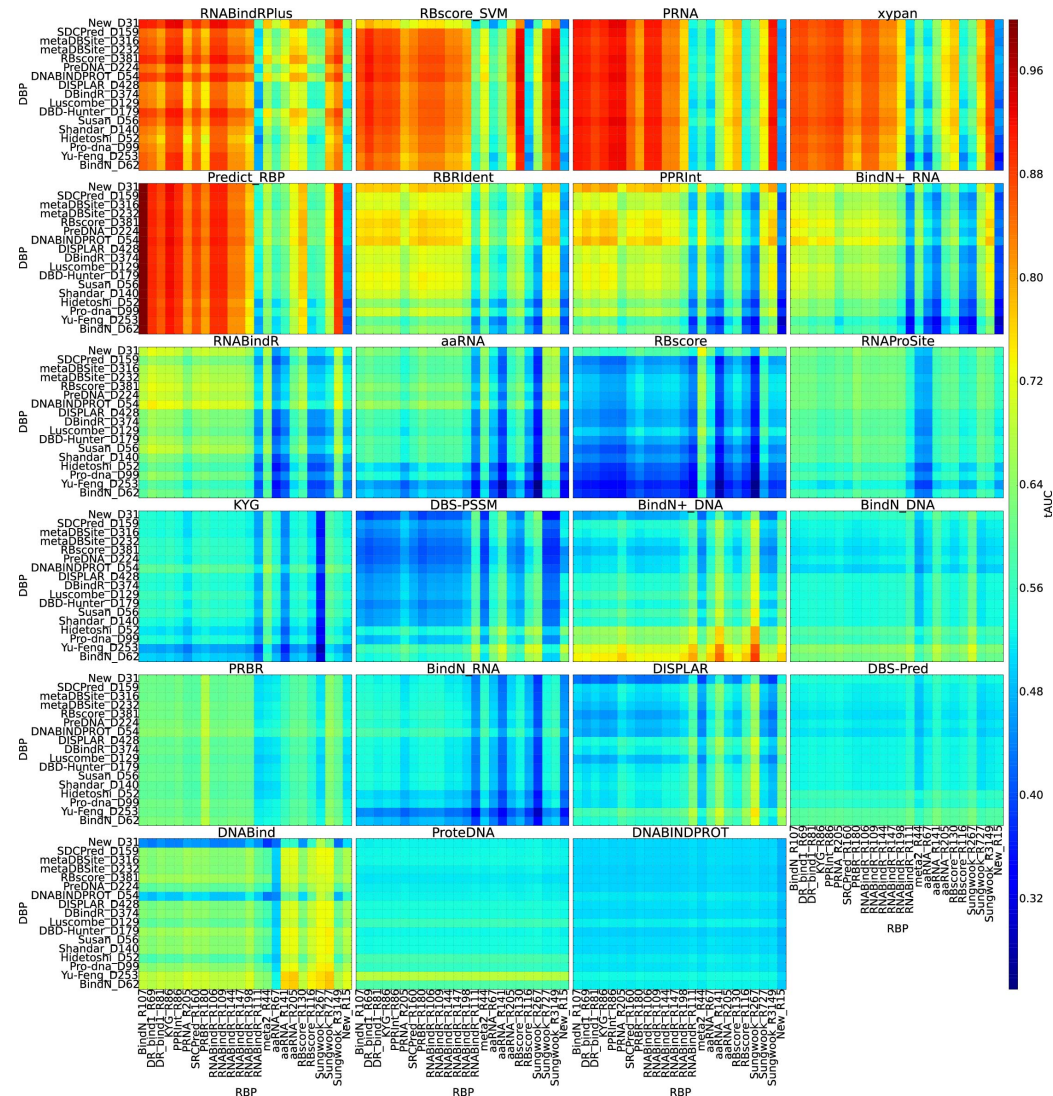
Bottom line of the prediction

Regardless of data set, regardless of binding site definition



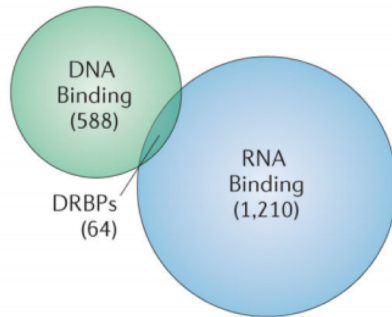
Distinguish DNA- from RNA-binding ?

No program
can correctly
distinguish all
data sets!

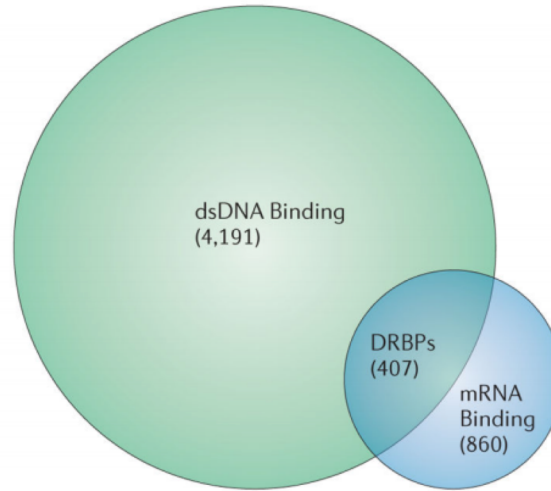


Why not?

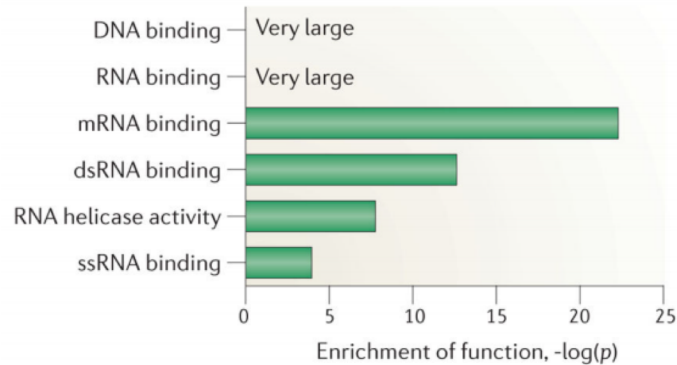
a Gene ontology (QuickGO)



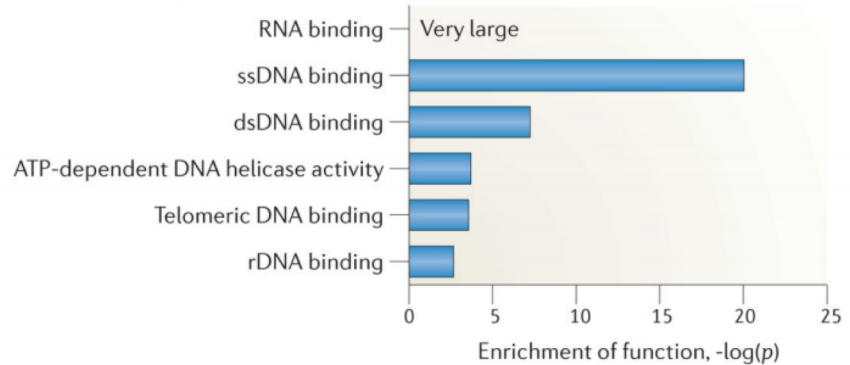
b Global nucleic acid interaction assays



c Gene ontology of DNA binding proteins

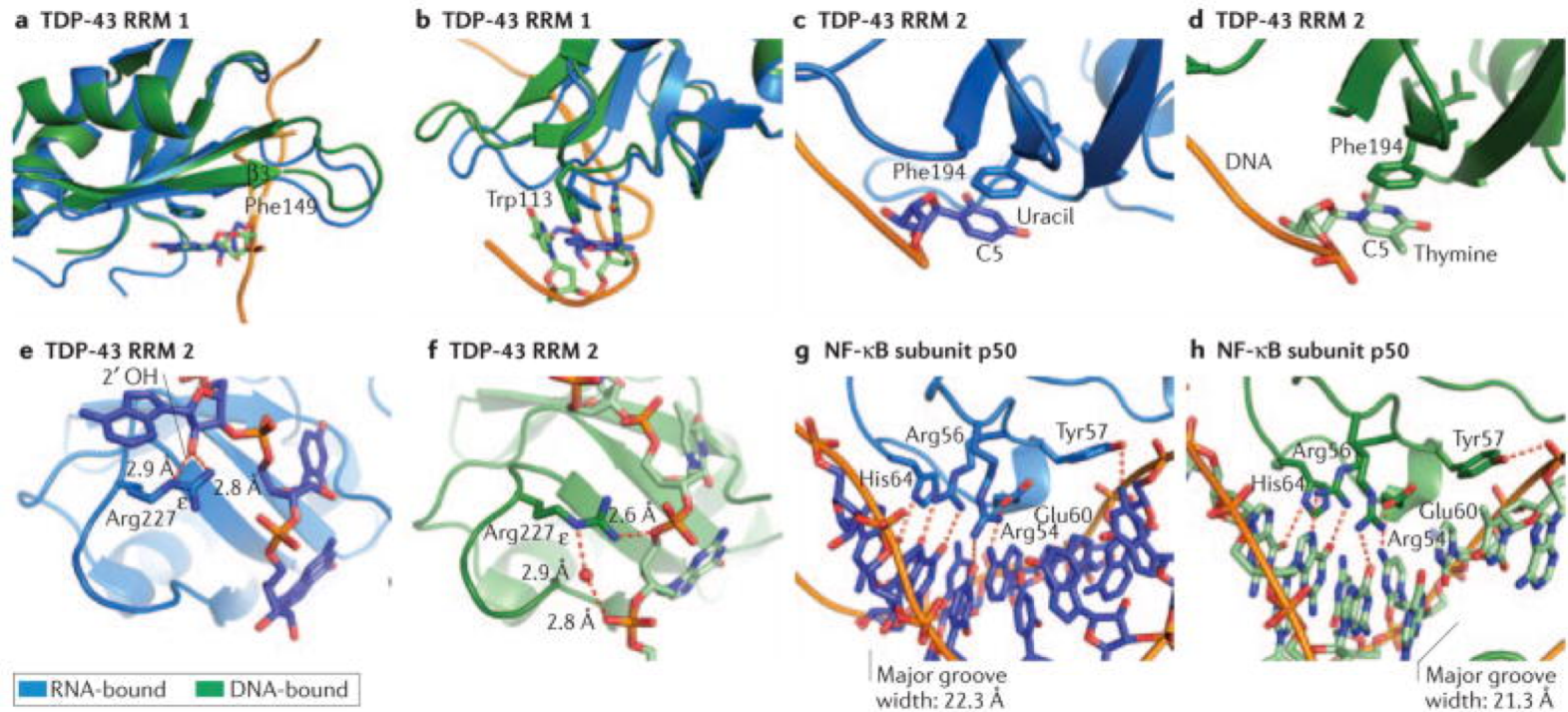


d Gene ontology of RNA binding proteins



Nat Rev Mol Cell Biol. 2014
November ; 15(11): 749–760

DRBP examples



- DRBP may bind DNA and RNA at the same position.

Hints from RBscore

- Energy funnels on surface.
- Train on RBP, predictive on DBP
 - RNA-binding and DNA-binding proteins may share similar rule!
- Stabler in accuracy (regardless dataset bias, distance cutoff bias)
- RBscore predicts ligand binding regions.
 - Can the driving force be identical?
- References
 - *Nucleic Acids Res*, 43(11):5340-51.
 - *PLoS Comput Biol* 11(12): e1004639.

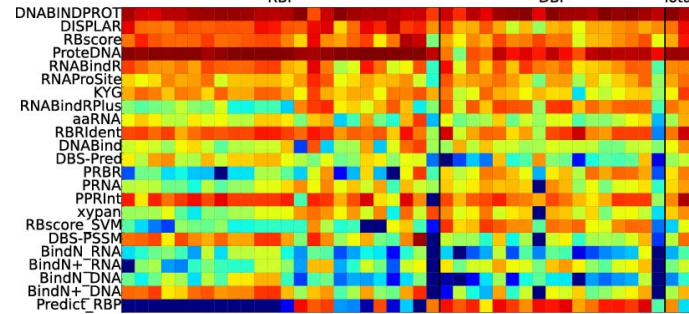
Acknowledgements

- Eric Westhof
- Alvis Brazma
- Sarah Teichmann

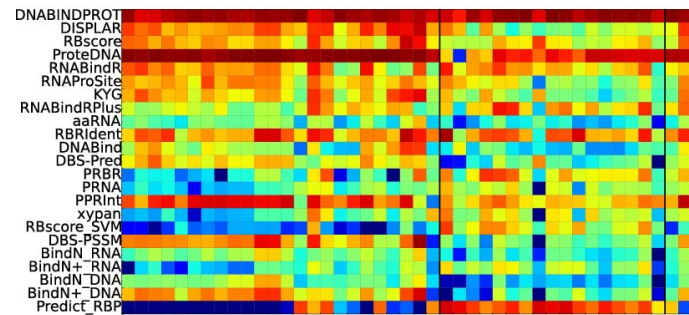
Accuracies are not very stable

MAVR ($\max(\frac{\Delta wAUC}{\Delta d}) \times 200$)

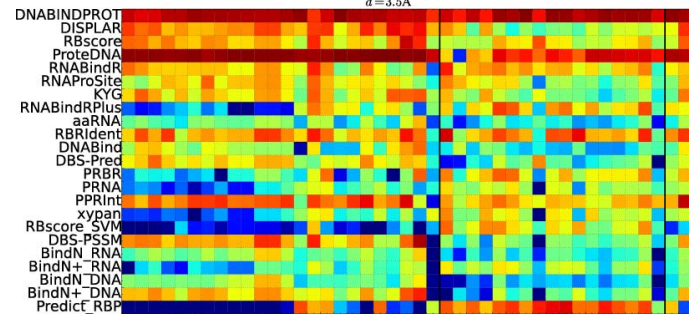
RBP DBP Total



MAV ($\max(\Delta wAUC) \times 100$)

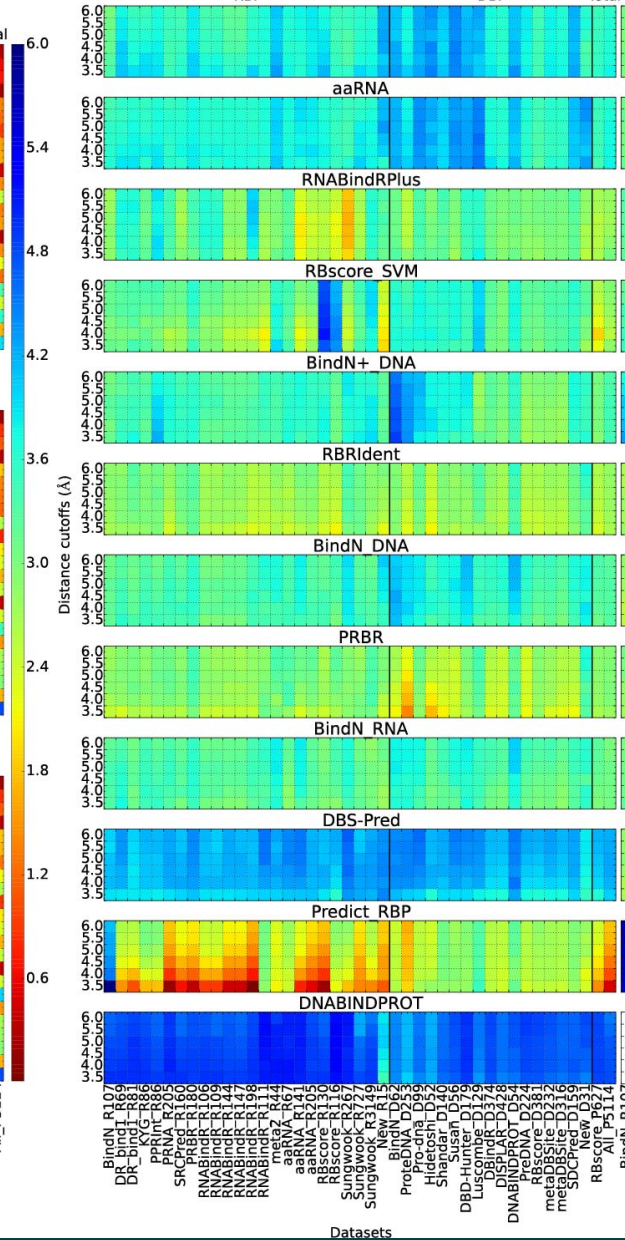


CAVR ($\sum_{d=3.5\text{\AA}}^{6.0\text{\AA}} \frac{\Delta wAUC}{\Delta d} \times 100$)



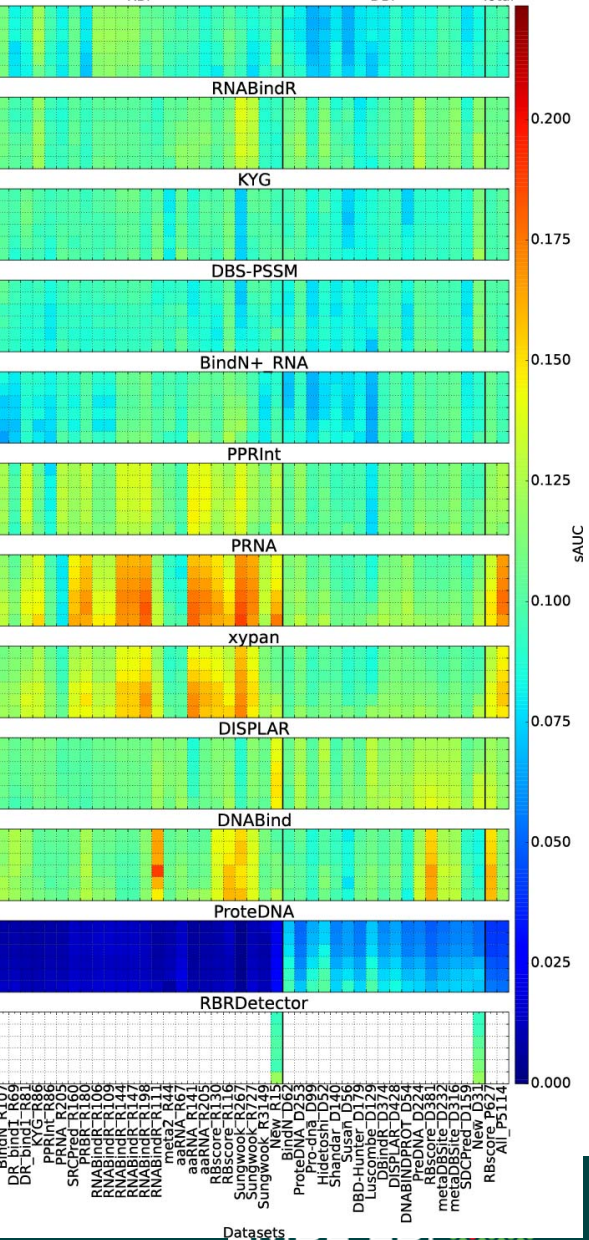
Datasets

RBP RBscore DBP Total



Datasets

RBP RNAProSite DBP Total



Datasets



17 Prediction methods

RNA																	
Sequence-based																	
Year	Name	Server	Program	Website	Binding Site Definition	PSSM	RP	ASA	HP	SS	EC	Q	SA	Training	Window	Dataset	reference
2006	BindN	✓		http://bioinfo.ggc.org/bindn/	3.5Å				✓			✓		SVM	11	R107(PRINR25),D62(PDNA-62)	[8]
2007	RNABindR	✓		http://einstein.cs.iastate.edu/RNABindR/	5Å	✓								NB	25	R147	[20]
2008	PPRInt	✓		http://www.imtech.res.in/raghava/pprint/	6Å	✓	✓							SVM	17	R107 from BindN, R86 from PPRInt	[23]
2008	RNAproB	✓			6Å/3.5Å	✓								SVM	25	R107 from BindN, R86 from PPRInt, R109 from RNABindR	[35]
2008	PRINTR	✓		http://210.42.106.80/printr/	ENTANGLE	✓								SVM	15	R109 from RNABindR	[16]
2008	RISP	✓		http://grc.seu.edu.cn/RISP	3.5Å	✓								SVM	7	R147 from RNABindR, R71(PRNA-71,NA)	[36]
2009	PIRaNha	✓		http://www.bioinformatics.sussex.ac.uk/PIRANHA	3.9Å	✓	✓	✓	✓					SVM	23	R81 from PPRInt,R42(NA)	[37]
2010	BindN+	✓		http://bioinfo.ggc.org/bindn+/	3.5Å	✓	✓		✓			✓		SVM	11	R107(PRINR25),D62(PDNA-62)	[9]
2010	NAPS	✓		http://proteomics.bioengr.uic.edu/NAPS/	4.5Å	✓								DT	7	R109 from RNABindR, D84 from Pro-dna, D274 from DISIS, D62(PDNA-62)	[7]
2010	PRBR	✓		http://www.cbi.seu.edu.cn/PRBR/optional.php	3.5Å	✓	✓							RF	11	R180 (RBP-180) R3149(PRI3149),R727(PRI727),R267(PRI267)	[22]
2011	Sungwook				H-bond			✓	✓			✓		SVM	9	R107 from BindN, R86 from PPRInt, R109 from RNABindR	[38]
2011	SRCPred	✓		http://tardis.nibio.go.jp/netasa/srcpred/	3.5Å	✓								NN	5	R160(PRNA160)	[39]
2011	Predict_RBP	✓		http://cic.scu.edu.cn/bioinformatics/Predict_RBP.rar	ENTANGLE	✓	✓	✓						SVM	15	R107 from BindN, R86 from PPRInt, R109 from RNABindR	[17]
Meta-server																	
2011	meta2	✓		https://genesilico.pl/meta2	3.5Å											R44, R38 from OPRA, R180 from PRBR, R111 from RNABindR, R81 from PIRaNha, R86 from KYG	[4]
2012	Qian-Zhong				6Å/3.5Å/ENTANGLE	✓		✓						SVM	25	R107 from BindN, R86 from PPRInt, R109 from RNABindR	[40]
2014	RNABindRPlus	✓		http://einstein.cs.iastate.edu/RNABindRPlus/	5Å	✓								SVM	21	R28,R44,R111,R198	[21]
2015	RBRIdent	✓		http://166.111.152.91/RBRIdent	ENTANGLE	✓				✓		✓		RF	9	R281	[41]
Structure-based																	
2006	KYG	✓		http://cib.cf.ocha.ac.jp/KYG/	7Å	✓	✓							Function		R86	[14]
2008	RsiteDB	✓		http://bioinfo3d.cs.tau.ac.il/RsiteDB/	7Å								✓	Clustering			[15]
2008	DR_bind1	✓		http://drbind.limlab.ljms.sinica.edu.tw/	HBPLUS			✓			✓	✓		Function		D56 from Susan 2003, D69, R81 R147(R144) and R109 from RNABindR	[42]
2009	PRIP	✓		http://www.qfab.org/PRIP	5Å	✓		✓						SVM	19	RNABindR	[43]
2010	OPRA		By contact		4Å				statistical potentials					Function		R316,R38 R250(RB250),R212(RB212),RBD292(NA)	[44]
2011	DRNA	✓			4.5Å								✓	Function		R147 from RNABindR	[45]
2010	SStruct-NB				5Å		✓							NB		R147 from RNABindR	[46]
2010	PRNA	✓		http://doc.aporc.org/wiki/PRNA	ENTANGLE	✓	✓	✓	✓		✓			RF	5	R205	[32]
2014	aaRNA	✓		http://sysimm.ifrec.osaka-u.ac.jp/aarna/	3.5Å	✓		✓		✓				NN	11	R67,R141,R205	[27]
2014	RBRDetector	✓		http://lbi.hzau.edu.cn/rbrdetector	4.5Å, 10%ASA	✓								SVM	11	R264, R75	[29]
2014	Xiaoyong	✓			ENTANGLE	✓		✓		✓		✓		RF	5	R205(PRNA)	[31]
2015	RBScore	✓		http://ahsoka.u-strasbg.fr/rbscore/	3.5-6Å		✓	✓			✓	✓				R130,R116	[28]
2015	RNAProSite	✓		lilab.ecust.edu.cn/NABind/													
DNA																	
Sequence-based																	
2004	DBS-Pred	✓		http://www.abren.net/dbs-pred/	3.5Å		✓	✓		✓				NN	3	D62(PDNA-62), NRTF-915	[12]
2005	DBS-PSSM	✓		http://www.abren.net/dbs-ssm/	3.5Å				✓		✓	✓		NN	5	D62(PDNA-62), PDNA-RDN(NA), PDNA-NR90(NA)	[13]
2006	DNABindR				ΔASA-1	✓		✓		✓	✓			NB	9	D171	[47]
2007	DISIS	✓		http://cubic.bioc.columbia.edu/services/disis	6Å	✓		✓		✓				SVM	9	D274	[48]
2007	DP-Bind	✓		http://lcg.rit.albany.edu/tp-bind/	3.5Å	✓								kernel regression		D62(PDNA-62)	[49]
2009	ProteDNA	✓		http://protdna.csbb.ntu.edu.tw/	4.5Å	✓								SVM, SSEA	11	D253	[50]
2009	DbindR	✓		http://www.cbi.seu.edu.cn/DBindR/DBindR.htm	3.5Å	✓				✓				RF	11	D374	[51]
2009	SDCPred	✓		http://sdcpred.netasa.org/	3.5Å	✓	✓							NN	5	D159(PDNA159)	[52]
2014	Byungkyu				H-bond		✓							SVM	9	D143	[53]
Structure-based																	
1999	Hidetoshi				3.5Å									Function		D52	[54]
2003	Susan				ΔASA-1							✓		Patch analysis		D56	[55]
2005	DBS-kernel				4.5Å		✓					✓		SVM		D83(NA)	[56]
2005	Pro-dna	✓		bioinformatics.bioengr.uic.edu/pro-dna/	4.5Å			✓		✓				SVM		D99 (D96,D50)	[57]
2005	PreDs	✓		http://pre-s.protein.osaka-u.ac.jp/~preDs/	3.0Å									Function		D63	[58]
2007	DISPLAR	✓		http://pipe.scs.fsu.edu/displar.html	5Å	✓								NN	15	D428	[10]
2007	DR_bind1	✓		http://drbind.limlab.ljms.sinica.edu.tw/	HBPLUS									Function		D56 from Susan 2003, D69, R81	[26]
2008	DBD-Hunter	✓		http://cssb.biology.gatech.edu/skolnick/webserve/DBD-Hunter/	4.5Å								✓	Function		D179(DB179)	[59]
2010	DNABINDPROT	✓		http://www.prc.boun.edu.tr/appserv/prc/dnabindprot/	NUCLOT		✓				✓			GNM	3	D54	[24]
Meta-server																	
2011	metaDBSite	✓		http://projects.biotech.tu-dresden.de/metadbsite/	3.5Å									NA		D316(PDNA-316),D232(PDNA-232)	[34]
2011	Xiong				4.5Å, 10%ASA	✓								SVM	11	D206	[60]
2012	Sucharita				ΔASA-0.1						✓	✓		SVM		D130(NA)	[61]
2013	Duo-Duo				4.0Å	✓								SVM	11	D62(PDNA-62)	[62]