Supplementary Data for "CompAnnotate: A comparative approach to annotate base pairing interactions in RNA 3D structures."

Shahidul Islam, Ping Ge and Shaojie Zhang *

1 Relevant atoms for acceptors, donors

To make geometric assesment, different atoms from the bases are considered. We call them *relevant atoms*. By default, we have included C1′, C2 and C5 as *relevant atoms*, as they are available in all the bases, in similar positions. Additionally, we pick at most 2 pairs of closest *involved atoms* (defined in the next paragraph).

In the base pairing interactions, different atoms participate in the formation of hydrogen bonds. Which particular atoms are participating as the acceptors and the donors of a hydrogen bond, depend on the participating edges of nucleotides. As the structure data only contains heavy atoms information, the closest heavy ones to the participating atoms are considered as the *involved atoms* in CompAnnotate for the observed distance comparisons. From the interactions characteristics defined by Leontis et al. [1], a list of involved atoms for acceptors and donors has been extracted. For a set of well observed types of base pairs, participating atoms can be specified [1].

Those atoms have been used to define involved atoms in CompAnnotate analysis, whenever possible. Otherwise closest heavy involved atoms from the pair of bases for acceptors and donors has been considered. The list of all involved atoms for acceptors, donors for each bases are given in Table S1, and the corresponding edge-wise involved atom list for different bases are given in Table S2, S3, S4 and S5.

Table S1: Involved atoms in nucleotides

Base	Acceptors	Donors
A	N1, N3, N7	N6, C2, C8, O2′
G	N3, N7, O6	N1, N2, C8, O2′
C	N3, O2	N4, C5, C6, O2′
U	O2, O4	N3, C5, O2′

---

*To whom correspondence should be addressed. Tel: +1 407 8236095; Fax: +1 407 8235835; Email: shzhang@eecs.ucf.edu

Table S2: Involved atoms (Adenine)

Edge	Acceptors	Donors
W-C	N1	N6, C2
Hoogsteen	N7	N6, C8
Sugar	N1, N3	C2, O2′

Table S3: Involved atoms (Guanine)

Edge	Acceptors	Donors
W-C	O6	N1, N2
Hoogsteen	O6, N7	C8
Sugar	N3	N2, O2′

Table S4: Involved atoms (Cytosine)

Edge	Acceptors	Donors
W-C	O2, N3	N4
Hoogsteen		N4, C5, C6
Sugar	O2	O2′

Table S5: Involved atoms (Uracil)

Edge	Acceptors	Donors
W-C	O2, O4	N3
Hoogsteen	O4	C5
Sugar	O2	O2′

2 CompAnnotate performance without STAR3D alignment (using sequence-based alignment as the input)

Table S6 and S7 show the annotation sensitivity and precision when we replaced the STAR3D alignment with sequence based alignment in the CompAnnotate pipeline. It shows that CompAnnotate is not dependent on STAR3D to make improvement in base pairing annotation, even though it is obvious that STAR3D alignment gives the best possible outcome.

Table S6: The improvements in base-pairing annotation sensitivity for five tools by using CompAnnotate

Method	Base-pair Type	# of BP in Benchmark	Regular		CompAnnotate	
			# of BP Detected	%	# of BP Detected	%
MC-Annotate	Canonical	2578	2063	80.02	2158	**83.71**
	Non-Canonical	1224	334	27.29	714	**58.33**
RNAView	Canonical	2634	2249	85.38	2254	**85.57**
	Non-Canonical	2782	999	35.91	1492	**53.63**
FR3D	Canonical	2622	2343	**89.36**	2307	87.99
	Non-Canonical	2788	1382	49.57	1710	**61.33**
DSSR	Canonical	2618	2284	87.24	2287	**87.36**
	Non-Canonical	1640	668	40.73	976	**59.51**
ClaRNA	Canonical	2518	2085	82.80	2166	**86.02**
	Non-Canonical	1456	500	34.34	887	**60.92**

Table S7: The improvements in base-pairing annotation precision for five tools by using CompAnnotate

Method	Base-pair Type	# of BP in Benchmark	Regular		CompAnnotate	
			Match/ Conflict	%	Match/ Conflict	%
MC-Annotate	Canonical	2578	2063/51	97.59	2158/28	**98.72**
	Non-Canonical	1224	334/319	51.15	714/162	**81.51**
RNAView	Canonical	2634	2249/153	93.63	2254/65	**97.20**
	Non-Canonical	2782	999/1338	42.75	1492/569	**72.39**
FR3D	Canonical	2622	2343/160	93.61	2307/55	**97.67**
	Non-Canonical	2788	1382/1151	54.56	1710/471	**78.40**
DSSR	Canonical	2618	2284/120	95.01	2287/38	**98.37**
	Non-Canonical	1640	668/689	49.23	976/307	**76.07**
ClaRNA	Canonical	2518	2085/27	98.72	2166/23	**98.95**
	Non-Canonical	1456	500/203	71.12	887/118	**88.26**

3 Acceptance ratio of base-pairing inferability relations

Table S8, S9, S10 and S11 show the acceptance ratio for CompAnnotate (RNAView), CompAnnotate (FR3D), CompAnnotate (DSSR), CompAnnotate (ClaRNA).

Table S8: Acceptance ratios of different inferability relationship for RNAView annotation, using reference structure in PDB 3R8S for target structure in PDB 2B9N.

Target (Ref.)	Relationship	Total # of BP	Accepted # of BP	Accept %
Canonical	Common	673	673	100.00
	Inferable — Case 1	93	62	66.67
	Inferable — Case 2	49	1	2.04
	Non-Inferable — Case 1	57	0	0.00
	Non-Inferable — Case 2	17	17	100.00
Non-Canonical	Common	264	264	100.00
	Inferable — Case 1	634	341	53.79
	Inferable — Case 2	387	4	1.03
	Non-Inferable — Case 1	58	0	0.00
	Non-Inferable — Case 2	14	14	100.00

Table S9: Acceptance ratios of different inferability relationship for FR3D annotation, using reference structure in PDB 3R8S for target structure in PDB 2B9N.

Target (Ref.)	Relationship	Total # of BP	Accepted # of BP	Accept %
Canonical	Common	696	696	100.00
	Inferable — Case 1	70	40	57.14
	Inferable — Case 2	44	1	2.27
	Non-Inferable — Case 1	56	0	0.00
	Non-Inferable — Case 2	35	35	100.00
Non-Canonical	Common	384	384	100.00
	Inferable — Case 1	544	263	48.35
	Inferable — Case 2	367	2	0.54
	Non-Inferable — Case 1	68	0	0.00
	Non-Inferable — Case 2	25	25	100.00

Table S10: Acceptance ratios of different inferability relationship for DSSR annotation, using reference structure in PDB 3R8S for target structure in PDB 2B9N.

Target (Ref.)	Relationship	Total # of BP	Accepted # of BP	Accept %
Canonical	Common	684	684	100.00
	Inferable — Case 1	90	51	56.67
	Inferable — Case 2	44	1	2.27
	Non-Inferable — Case 1	56	0	0.00
	Non-Inferable — Case 2	21	21	100.00
Non-Canonical	Common	177	177	100.00
	Inferable — Case 1	382	203	53.14
	Inferable — Case 2	199	3	1.51
	Non-Inferable — Case 1	38	0	0.00
	Non-Inferable — Case 2	6	6	100.00

Table S11: Acceptance ratios of different inferability relationship for ClaRNA annotation, using reference structure in PDB 3R8S for target structure in PDB 2B9N.

Target (Ref.)	Relationship	Total # of BP	Accepted # of BP	Accept %
Canonical	Common	626	626	100.00
	Inferable — Case 1	124	81	65.32
	Inferable — Case 2	14	0	0.00
	Non-Inferable — Case 1	48	0	0.00
	Non-Inferable — Case 2	10	10	100.00
Non-Canonical	Common	113	113	100.00
	Inferable — Case 1	368	206	55.98
	Inferable — Case 2	72	0	0.00
	Non-Inferable — Case 1	24	0	0.00
	Non-Inferable — Case 2	3	3	100.00

4 Broke-down performance details of all data set for each method

The sensitivity and precision of the four datasets for each considered methods are given in Table S12, S13, S14, S15, S16, S16, S17, S18, S19, S20, S21. Here we can see how the annotation of a particular target RNA is improved.

Table S12: sensitivity compared for MC-Annotate annotation

Method	Base-pair Type	# of BP in Benchmark	Regular		CompAnnotate	
			# of BP Detected	%	# of BP Detected	%
2B9N (3R8S)	Canonical	842	622	73.87	683	**81.12**
	Non-Canonical	405	91	22.47	233	**57.53**
2B9N (2ZJR)	Canonical	842	622	73.87	691	**82.07**
	Non-Canonical	405	91	22.47	205	**50.62**
1YL4 (4GD1)	Canonical	447	402	**89.93**	400	89.49
	Non-Canonical	207	69	33.33	127	**61.35**
2B9M (4GD1)	Canonical	447	409	**91.50**	406	90.83
	Non-Canonical	207	87	42.03	137	**66.18**

Table S13: precision compared for MC-Annotate annotation

Method	Base-pair Type	# of BP in Benchmark	Regular		CompAnnotate	
			Match/ Conflict	%	Match/ Conflict	%
2B9N (3R8S)	Canonical	842	622/15	97.65	683/7	**98.99**
	Non-Canonical	405	91/108	45.73	233/42	**84.73**
2B9N (2ZJR)	Canonical	842	622/15	97.65	691/11	**98.43**
	Non-Canonical	405	91/108	45.73	205/50	**80.39**
1YL4 (4GD1)	Canonical	447	402/11	97.34	400/6	**98.52**
	Non-Canonical	207	69/50	57.98	127/29	**81.41**
2B9M (4GD1)	Canonical	447	409/10	97.61	406/7	**98.31**
	Non-Canonical	207	87/37	70.16	137/33	**80.59**

Table S14: sensitivity compared for RNAView annotation

Method	Base-pair Type	# of BP in Benchmark	Regular		CompAnnotate	
			# of BP Detected	%	# of BP Detected	%
2B9N (3R8S)	Canonical	854	689	80.68	709	**83.02**
	Non-Canonical	917	270	29.44	467	**50.93**
2B9N (2ZJR)	Canonical	854	689	80.68	715	**83.72**
	Non-Canonical	917	270	29.44	452	**49.29**
1YL4 (4GD1)	Canonical	463	434	**93.74**	425	91.79
	Non-Canonical	474	232	48.95	268	**56.54**
2B9M (4GD1)	Canonical	463	430	**92.87**	425	91.79
	Non-Canonical	474	237	50.00	279	**58.86**

Table S15: precision compared for RNAView annotation

Method	Base-pair Type	# of BP in Benchmark	Regular		CompAnnotate	
			Match/ Conflict	%	Match/ Conflict	%
2B9N (3R8S)	Canonical	854	689/48	93.49	709/22	**96.99**
	Non-Canonical	917	270/418	39.24	467/133	**77.83**
2B9N (2ZJR)	Canonical	854	689/48	93.49	715/22	**97.01**
	Non-Canonical	917	270/418	39.24	452/213	**67.97**
1YL4 (4GD1)	Canonical	463	434/28	93.94	425/14	**96.81**
	Non-Canonical	474	232/238	49.36	268/104	**72.04**
2B9M (4GD1)	Canonical	463	430/31	93.28	425/15	**96.59**
	Non-Canonical	474	237/212	52.78	279/110	**71.72**

Table S16: sensitivity compared for FR3D annotation

Method	Base-pair Type	# of BP in Benchmark	Regular		CompAnnotate	
			# of BP Detected	%	# of BP Detected	%
2B9N (3R8S)	Canonical	857	719	83.90	724	**84.48**
	Non-Canonical	916	391	42.69	523	**57.10**
2B9N (2ZJR)	Canonical	857	719	83.90	736	**85.88**
	Non-Canonical	916	391	42.69	525	**57.31**
1YL4 (4GD1)	Canonical	454	442	**97.36**	436	96.04
	Non-Canonical	478	278	58.16	311	**65.06**
2B9M (4GD1)	Canonical	454	442	**97.36**	437	96.26
	Non-Canonical	478	330	69.04	332	**69.46**

Table S17: precision compared for FR3D annotation

Method	Base-pair Type	# of BP in Benchmark	Regular		CompAnnotate	
			Match/ Conflict	%	Match/ Conflict	%
2B9N (3R8S)	Canonical	857	719/54	93.01	724/16	**97.84**
	Non-Canonical	916	391/381	50.65	523/120	**81.34**
2B9N (2ZJR)	Canonical	857	719/54	93.01	736/17	**97.74**
	Non-Canonical	916	391/381	50.65	525/180	**74.47**
1YL4 (4GD1)	Canonical	454	442/25	94.65	436/13	**97.10**
	Non-Canonical	478	278/203	57.80	311/77	**80.15**
2B9M (4GD1)	Canonical	454	442/24	94.85	437/12	**97.33**
	Non-Canonical	478	330/148	69.04	332/82	**80.19**

Table S18: sensitivity compared for DSSR annotation

Method	Base-pair Type	# of BP in Benchmark	Regular		CompAnnotate	
			# of BP Detected	%	# of BP Detected	%
2B9N (3R8S)	Canonical	854	695	81.38	713	**83.49**
	Non-Canonical	532	179	33.65	303	**56.95**
2B9N (2ZJR)	Canonical	854	695	81.38	720	**84.31**
	Non-Canonical	532	179	33.65	297	**55.83**
1YL4 (4GD1)	Canonical	455	441	**96.92**	431	94.73
	Non-Canonical	288	141	48.96	174	**60.42**
2B9M (4GD1)	Canonical	455	438	**96.26**	429	94.29
	Non-Canonical	288	171	59.38	187	**64.93**

Table S19: precision compared for DSSR annotation

Method	Base-pair Type	# of BP in Benchmark	Regular		CompAnnotate	
			Match/ Conflict	%	Match/ Conflict	%
2B9N (3R8S)	Canonical	854	695/37	94.95	713/8	**98.89**
	Non-Canonical	532	179/214	45.55	303/83	**78.50**
2B9N (2ZJR)	Canonical	854	695/37	94.95	720/11	**98.50**
	Non-Canonical	532	179/214	45.55	297/110	**72.97**
1YL4 (4GD1)	Canonical	455	441/19	95.87	431/12	**97.29**
	Non-Canonical	288	141/130	52.03	174/60	**74.36**
2B9M (4GD1)	Canonical	455	438/21	95.42	429/13	**97.06**
	Non-Canonical	288	171/107	61.51	187/60	**75.71**

Table S20: sensitivity compared for ClaRNA annotation

Method	Base-pair Type	# of BP in Benchmark	Regular		CompAnnotate	
			# of BP Detected	%	# of BP Detected	%
2B9N (3R8S)	Canonical	818	630	77.02	683	**83.50**
	Non-Canonical	479	127	26.51	280	**58.46**
2B9N (2ZJR)	Canonical	818	630	77.02	687	**83.99**
	Non-Canonical	479	127	26.51	258	**53.86**
1YL4 (4GD1)	Canonical	441	400	90.70	402	**91.16**
	Non-Canonical	249	102	40.96	158	**63.45**
2B9M (4GD1)	Canonical	441	419	**95.01**	411	93.20
	Non-Canonical	249	144	57.83	169	**67.87**

Table S21: precision compared for ClaRNA annotation

Method	Base-pair Type	# of BP in Benchmark	Regular		CompAnnotate	
			Match/ Conflict	%	Match/ Conflict	%
2B9N (3R8S)	Canonical	818	630/5	**99.21**	683/8	98.84
	Non-Canonical	479	127/56	69.40	280/38	**88.05**
2B9N (2ZJR)	Canonical	818	630/5	**99.21**	687/8	98.85
	Non-Canonical	479	127/56	69.40	258/40	**86.58**
1YL4 (4GD1)	Canonical	441	400/10	97.56	402/6	**98.53**
	Non-Canonical	249	102/41	71.33	158/22	**87.78**
2B9M (4GD1)	Canonical	441	419/11	97.44	411/7	**98.33**
	Non-Canonical	249	144/40	78.26	169/24	**87.56**

References

[1] Leontis, N. B., Stombaugh, J., and Westhof, E. (Aug, 2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.,* **30**(16), 3497–3531.