

# lociPARSE: a locality-aware invariant point attention model for scoring RNA 3D structures

Sumit Tarafder and Debswapan Bhattacharya

Department of Computer Science, Virginia Tech, USA

## Background

- Knowledge-based statistical potentials and deep-learning for RNA quality estimation
- Deep-learning based methods rely on superposition-dependent RMSD
- Scoring should be invariant to relative motion among structural motifs

## Novelty

- First method targeting IDDT as ground truth
- Locality-aware invariant point attention based on atomic coordinates
- First deep learning method to estimate RNA local and global quality in a superposition free way

## Workflow

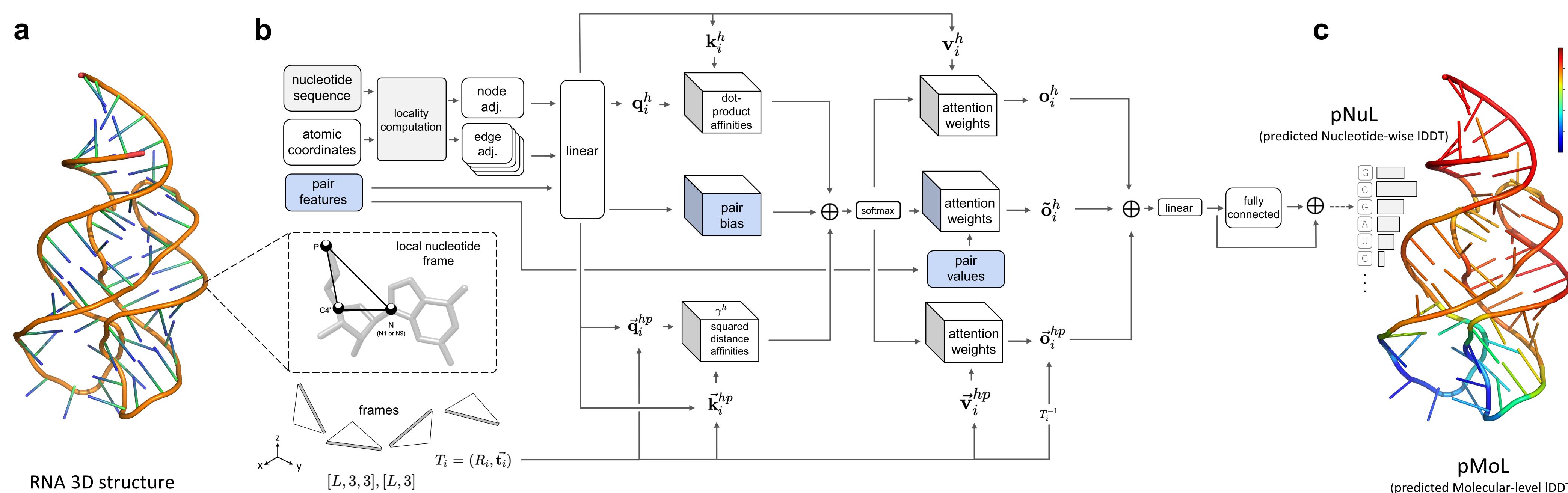
- Extract features from any input RNA 3D structure
- Generate per nucleotide frame from P, C4' and N coordinates
- Aggregate features based on K-nearest neighbors
- Compute invariant point attention operating on frames
- Predict quality score (0-1)
  - pMoL = predicted molecular-level IDDT
  - pNuL = predicted nucleotide-wise IDDT

## Source

<https://github.com/Bhattacharya-Lab/lociPARSE>



## Overview



## Performance on independent test set

**Table 1.** Performance on 30 independent RNA targets in terms of IDDT as ground truth, sorted in decreasing order of global Pearson's  $r$ . Best values are highlighted in bold.

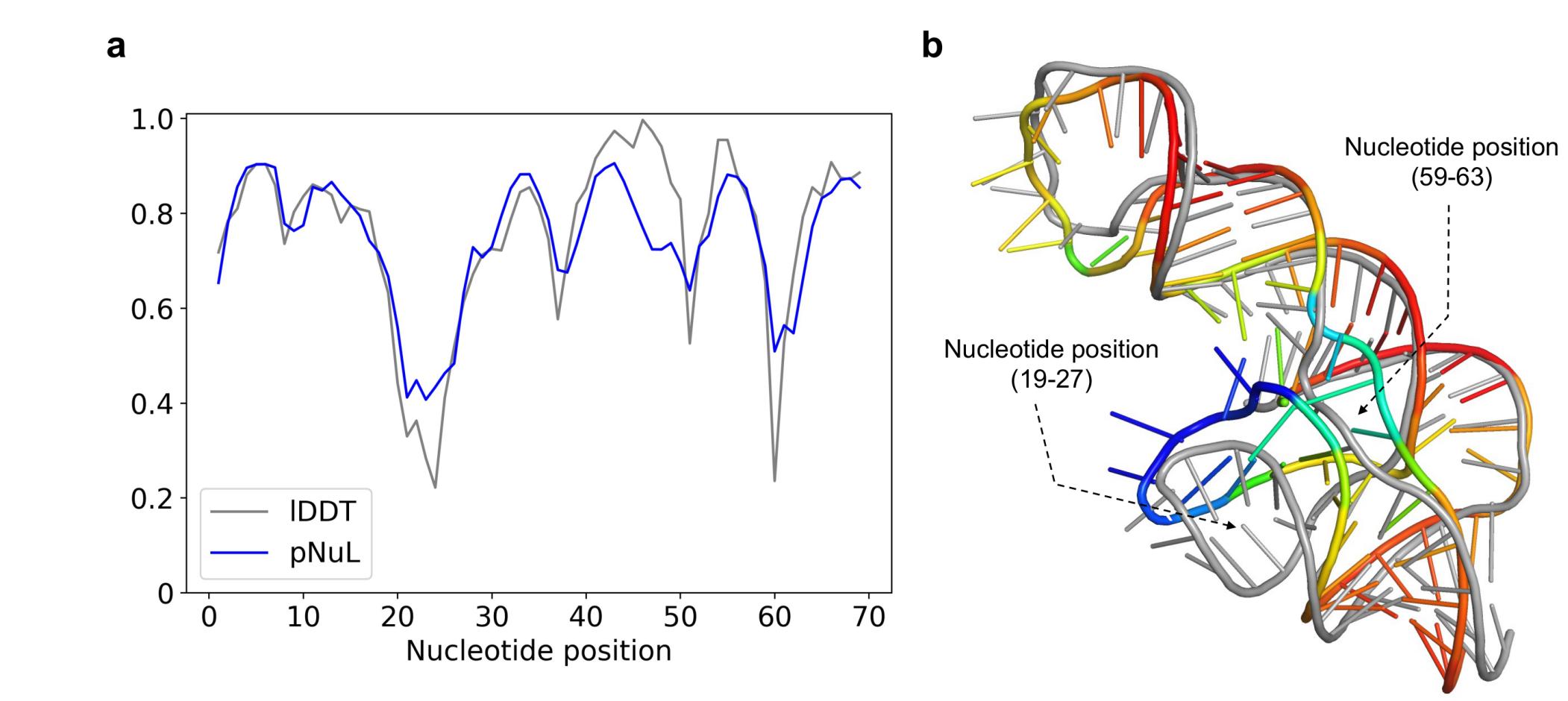
Method	Global					Per-target average			
	$r \uparrow$	$\rho \uparrow$	$\tau \uparrow$	Diff $\downarrow$	ROC $\uparrow$	$r \uparrow$	$\rho \uparrow$	$\tau \uparrow$	Loss $\downarrow$
lociPARSE	<b>0.67</b>	<b>0.71</b>	<b>0.55</b>	<b>0.06</b>	<b>0.91</b>	<b>0.77</b>	<b>0.75</b>	<b>0.64</b>	0.06
rsRNAsP	0.5	0.5	0.36	0.11	0.79	0.75	0.69	0.55	0.06
RASP	0.46	0.5	0.36	0.16	0.78	0.72	0.68	0.58	0.08
ARES	0.43	0.48	0.34	0.55	0.8	0.72	0.7	0.58	0.06
DFIRE-RNA	0.33	0.32	0.22	0.19	0.69	0.75	0.7	0.56	<b>0.05</b>
cgRNAsP	0.27	0.19	0.13	0.15	0.60	0.12	0.07	0.07	0.07
RNA3DCNN	0.19	0.17	0.12	0.64	0.58	0.47	0.36	0.27	0.06

## Performance on CASP15

**Table 2.** Performance on CASP15 RNA targets in terms of IDDT as ground truth, sorted in decreasing order of global Pearson's  $r$ . Best values are highlighted in bold.

Method	Global					Per-target average			
	$r \uparrow$	$\rho \uparrow$	$\tau \uparrow$	Diff $\downarrow$	ROC $\uparrow$	$r \uparrow$	$\rho \uparrow$	$\tau \uparrow$	Loss $\downarrow$
lociPARSE	<b>0.74</b>	<b>0.72</b>	<b>0.55</b>	0.12	<b>0.96</b>	<b>0.73</b>	<b>0.66</b>	<b>0.52</b>	<b>0.07</b>
ARES	0.33	0.35	0.25	0.27	0.78	0.55	0.53	0.42	0.13
RASP	0.33	0.42	0.3	0.17	0.80	0.55	0.54	0.42	0.17
rsRNAsP	0.31	0.41	0.3	<b>0.11</b>	0.83	0.66	0.61	0.47	0.11
DFIRE-RNA	0.29	0.3	0.22	0.39	0.77	0.69	0.6	0.46	0.16
cgRNAsP	0.29	0.32	0.24	0.12	0.80	0.59	0.5	0.37	0.11
RNA3DCNN	0	-0.01	-0.01	0.52	0.35	0.5	0.42	0.31	0.09

## Case study



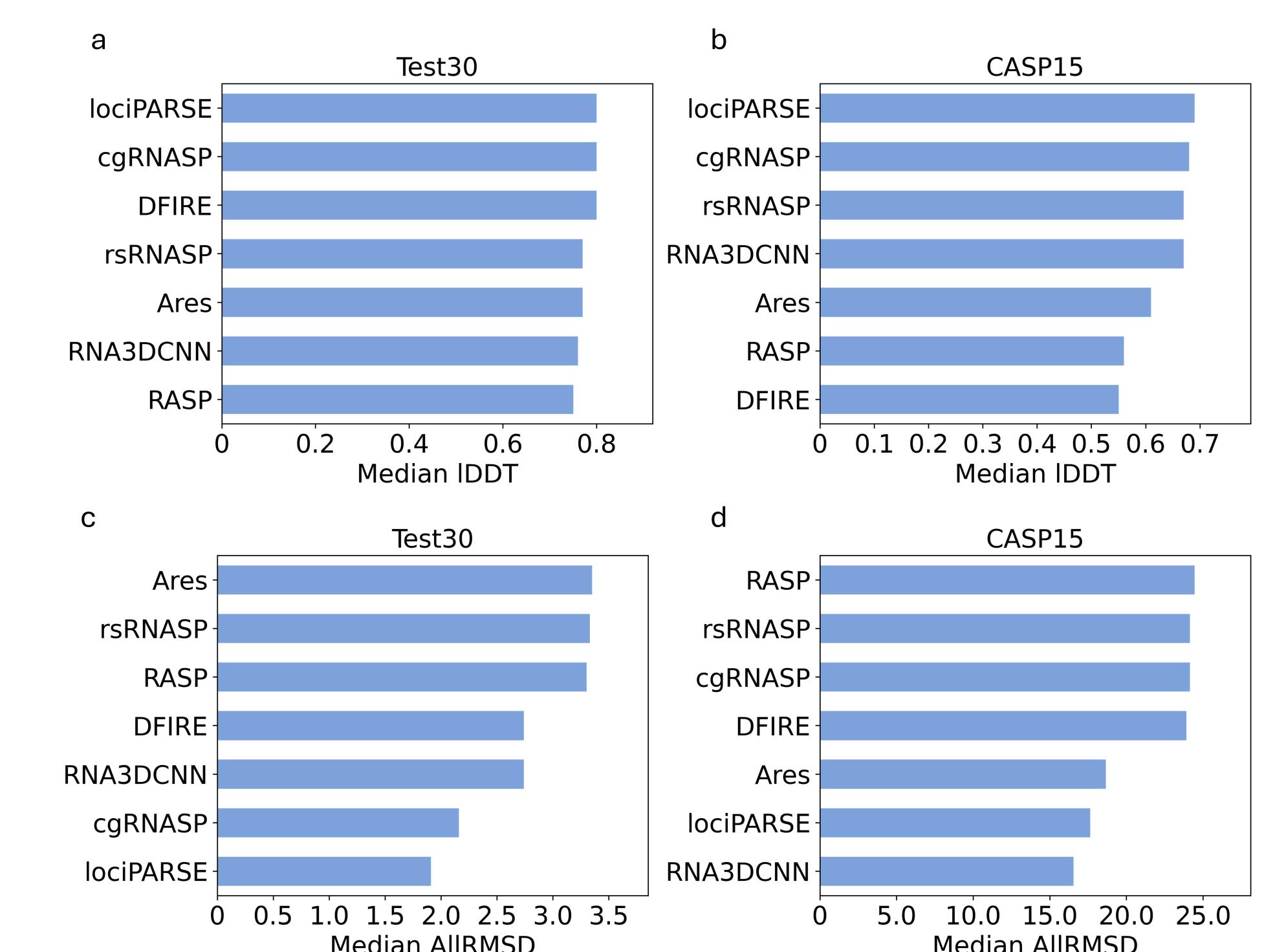
Per nucleotide prediction accuracy on 7QR3

## Local quality prediction

**Table 3.** Local nucleotide-wise scoring performance on CASP15 set. Best values are highlighted in bold.

Method	Local nucleotide-wise			
	$r \uparrow$	$\rho \uparrow$	$\tau \uparrow$	Diff $\downarrow$
lociPARSE	<b>0.72</b>	<b>0.73</b>	<b>0.53</b>	0.15
RNA3DCNN	0.17	0.19	0.13	0.24

## Ranking ability



Median scores of top ranked structure

## Conclusion

- Can be extended for RMSD or INF
- Multi-reference IDDT for enhanced expressivity