

Training stage	Name on HuggingFace	Source	Method	Number of sequences	Families	length [10, 199]	length [200, 499]	length [500, 999]	length [1000, 1999]	length [2000, inf]
Pre-training	rnacentral_synthetic	Sequences from RNA central	RNAstructure	226,729	All known families	176,486	49,463	780	0	0
Pre-training	ribo500-blast	Ribonanza Competition	RNAstructure + DMS and/or SHAPE	46,060	Unlabelled	46,049	11	0	0	0
Pre-training	bpRNA-1m	bpRNA-1m	Covariance analysis	66,715	Unlabelled, sRNA, tRNA	48,090	6,167	2,829	9,260	369
Fine-tuning	pri_miRNA	This work	RNAstructure + DMS	1,098	pri-miRNA	0	1,098	0	0	0
Fine-tuning	human_mRNA	This work	RNAstructure + DMS	1,456	mRNA	0	493	882	81	0
Testing	PDB	PDB	NMR, crystallography	355	Short non-coding RNA	343	6	6	0	0
Testing	viral_fragments	Peer-reviewed literature	RNAstructure + DMS	40	Viral RNA	12	17	11	0	0
Testing	lncRNA	Bugnon and al, 2022	RNAstructure + DMS	30	Long non-coding RNA	0	2	1	27	0
Testing	archiveII	Archive II	Covariance analysis	3,370	rRNA, tRNA, tmRNA, unlabelled	2,004	1,276	79	11	0