

Training stage	Name on HuggingFace	Source	Method	Number of sequences	Families	10 - 199 nt	200 - 499 nt	500 - 999 nt	1000 - 1999 nt	> 2000 nt
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	356	Short non-coding RNA	343	6	6	1	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	10	Long non-coding RNA	0	2	1	7	0
Testing	archivell_filtered	ArchiveII	Covariance analysis	355	rRNA, tRNA, tmRNA, unlabelled	242	65	43	5	0