

miRNA-mRNA interactions

白梓琳 2022/9/16

I Dataset

- PAR-CLIP and CLASH experiment datasets

1.1 Train

- site-level
 - positive: 33,142
 - negative: 32,284

	A	B	C	D	E	F
1	MIRNA_ID	ENSEMBL_GENE	GENE_SYMBOL	MIRNA_SEQ	MRNA_SEQ	LABEL
2	hsa-let-7a-5p	ENSG00000114573	ATP6V1A	UGAGGUAGUAGGU	GCCTGCTATTGAGG	1
3	hsa-let-7a-5p	ENSG00000104497	SNX16	UGAGGUAGUAGGU	ATTTGTTTAGTTTC	1
4	hsa-let-7a-5p	ENSG00000141682	PMAIP1	UGAGGUAGUAGGU	GCACATTGTATATC	1
5	hsa-let-7a-5p	ENSG00000174010	KLHL15	UGAGGUAGUAGGU	GAAGTTAGACACCT	1
6	hsa-let-7b-5p	ENSG00000130402	ACTN4	UGAGGUAGUAGGU	GGCCCTCATCTTCC	1
7	hsa-let-7c-5p	ENSG00000130402	ACTN4	UGAGGUAGUAGGU	GGAGGCCCTCATCT	1
8	hsa-let-7a-5p	ENSG00000130402	ACTN4	UGAGGUAGUAGGU	GGAGGCCCTCATCT	1
9	hsa-let-7d	ENSG00000130402	ACTN4	UGAGGUAGUAGGU	GGCCCTCATCTTCC	1
10	hsa-let-7b-5p	ENSG00000105202	FBL	UGAGGUAGUAGGU	ACCCCTTCCGCTCC	1
11	hsa-let-7c-5p	ENSG00000105202	FBL	UGAGGUAGUAGGU	ACCCCTTCCGCTCC	1
12	hsa-let-7a-5p	ENSG00000105202	FBL	UGAGGUAGUAGGU	ACCCCTTCCGCTCC	1
13	hsa-let-7d	ENSG00000105202	FBL	UGAGGUAGUAGGU	ACCCCTTCCGCTCC	1
14	hsa-let-7c-5p	ENSG00000170779	CDCA4	UGAGGUAGUAGGU	CGGTGAGCCGCAG	1
15	hsa-let-7a-5p	ENSG00000170779	CDCA4	UGAGGUAGUAGGU	CGGTGAGCCGCAG	1
16	hsa-let-7b-5p	ENSG00000164011	ZNF691	UGAGGUAGUAGGU	TATGTGCCCAAGTG	1

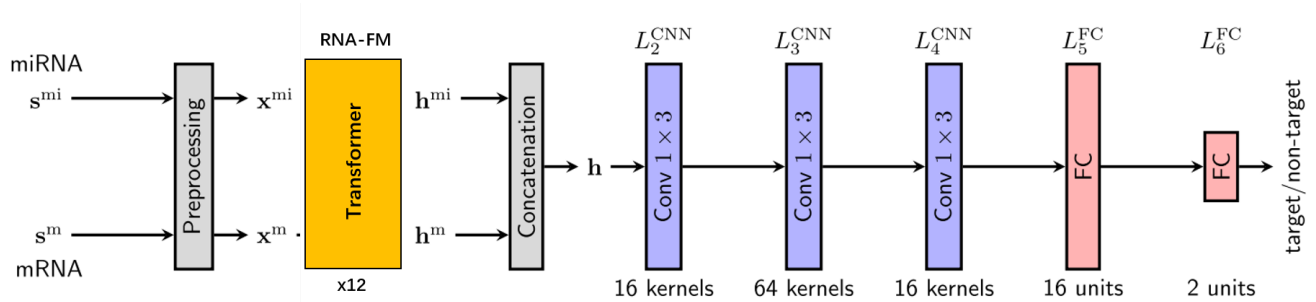
1.2 Test

- 10 test sets
- gene-level
 - positive: 151,956
 - negative: 548 (randomly sampled 548 positive pairs for ten times to balance the number of positive and negative pairs)

541	hsa-miR-99a-5p	NM_030791	0
542	hsa-miR-99a-5p	NM_031459	0
543	hsa-miR-99a-5p	NM_078467	0
544	hsa-miR-99a-5p	NM_153331	0
545	hsa-miR-99a-5p	NM_173595	0
546	hsa-miR-99a-5p	NM_181755	0
547	hsa-miR-99a-5p	NM_183078	0
548	hsa-miR-99a-5p	NM_183356	0
549	hsa-let-7a-5p	NM_032015	1
550	hsa-let-7b-3p	NM_001290257	1
551	hsa-let-7b-5p	NM_001009984	1
552	hsa-let-7b-5p	NM_001083589	1
553	hsa-let-7b-5p	NM_001145314	1
554	hsa-let-7b-5p	NM_001320923	1
555	hsa-let-7b-5p	NM_007055	1
556	hsa-let-7c-5p	NM_001130524	1

II Model

TargetFM



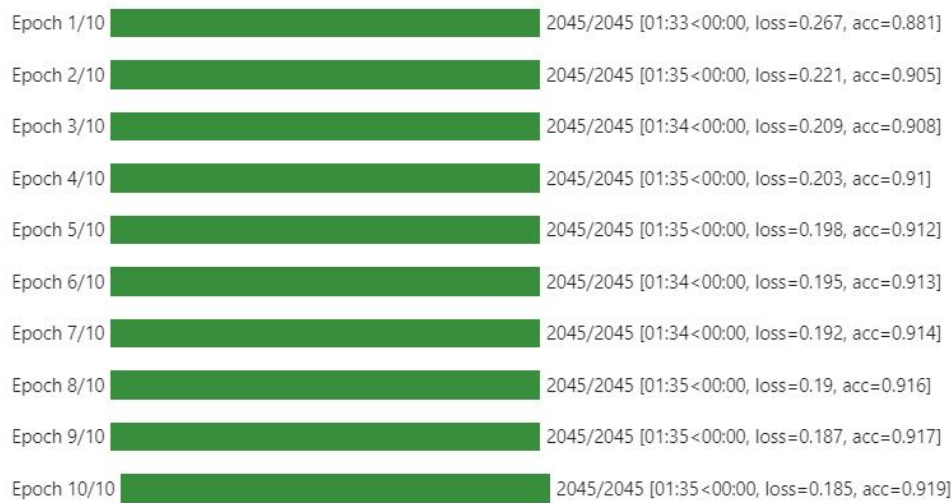
III Train & Inference

3.1 Train

Hyperparameter

- seed_match = 'offset-9-mer-m7'
- level = 'gene'

- batch_size = 32
- epochs = 10



3.2 Inference

Hyperparameter

- seed_match = 'offset-9-mer-m7'
- level = 'gene'

Result

DeepTarget

	accuracy
set 1	0.7729
set 2	0.7934
set 3	0.7948
set 4	0.7690
set 5	0.7707
set 6	0.7758
set 7	0.7787
set 8	0.7822
set 9	0.7753
set 10	0.7907
average	0.7804

Our model

156/156 [00:06<00:00]

4962
acc: 79.32285368802901 %

171/171 [00:07<00:00]

5457
acc: 84.91845336265347 %

162/162 [00:07<00:00]

5153
acc: 82.8449446924122 %

155/155 [00:06<00:00]

4956
acc: 81.79983857949959 %

160/160 [00:07<00:00]

5101
acc: 81.96432072142717 %

169/169 [00:07<00:00]

5388
acc: 79.28730512249443 %

160/160 [00:07<00:00]

5102
acc: 79.08663269306155 %

162/162 [00:07<00:00]

5167
acc: 82.11728275595122 %

156/156 [00:06<00:00]

4986
acc: 82.49097472924187 %

162/162 [00:07<00:00]

5172
acc: 83.33333333333334 %
avg = 81.71659396781038 %
