



**RBF-TSS: IDENTIFICATION OF TRANSCRIPTION START SITE
IN HUMAN USING RADIAL BASIS FUNCTIONS NETWORK AND
LOCALIZED OLIGONUCLEOTIDE FREQUENCIES**

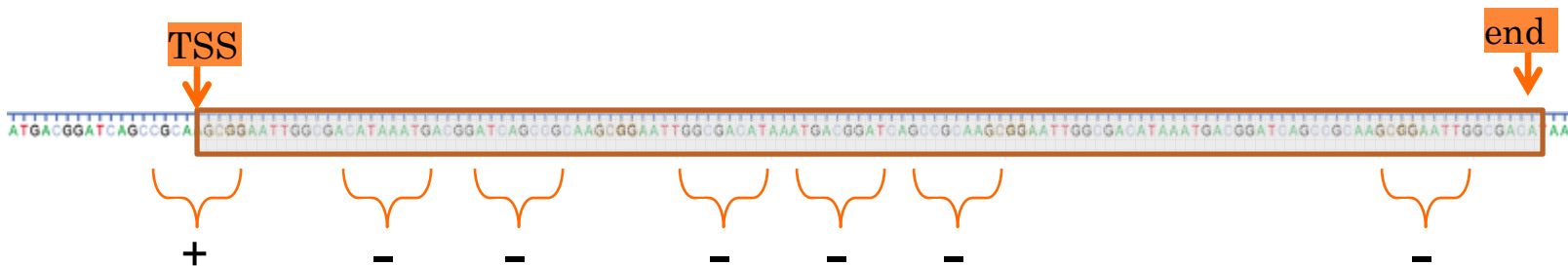
Rami N. Mahdi, Eric C. Rouchka

**Department of Computer Engineering and Computer Science,
University of Louisville, Louisville, Kentucky, United States of
America**

DATASET

○ Training & Validation

- 8542 genes were extracted from dbTSS version 4 and UCSC human genome sequence assembly and annotation version 16
- The TSS of every gene was taken as a positive sample
- 10 negative samples were randomly extracted from within each gene downstream from the TSS



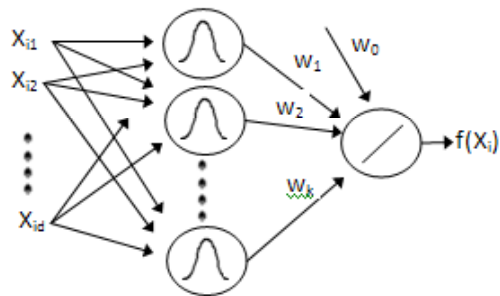
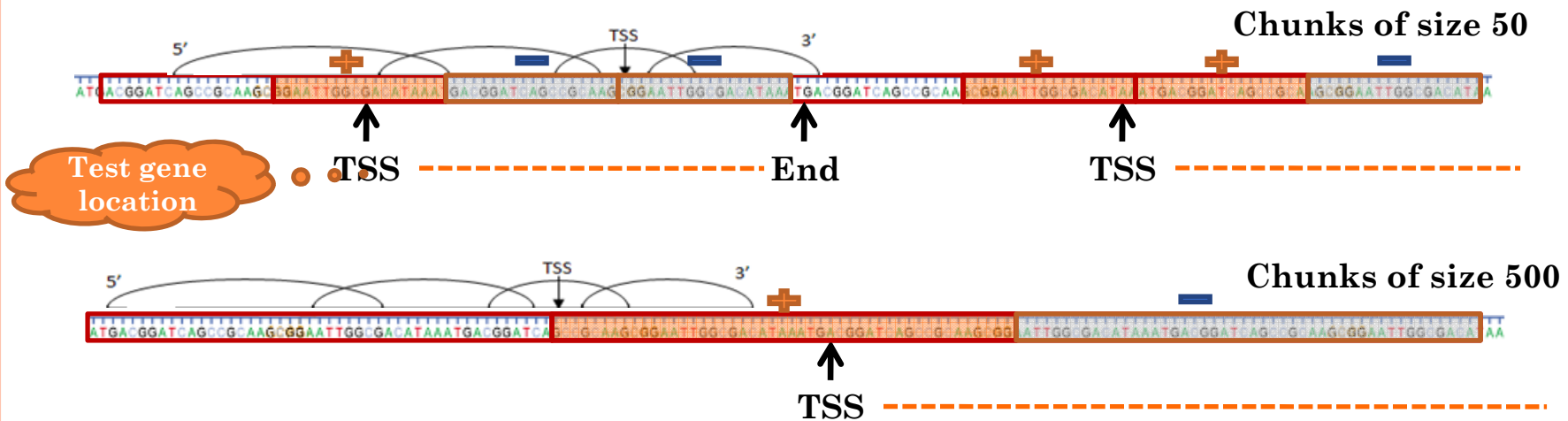
- 50% of the data was used for training and 50% for validation

○ Testing

- 1024 genes for testing were extracted from dbTSS version 5 and UCSC human genome sequence assembly and annotation version 17

TESTING

- Every chromosome strand is uniformly divided into non-overlapping chunks and only the chunks within test genes are labeled for testing.



Confidence score is computed at every base and the score of the whole chunk is the maximum of its bases

Chunk size	# of +	# of -
50	1588	1,087,666
500	943	108,782

True Labeling

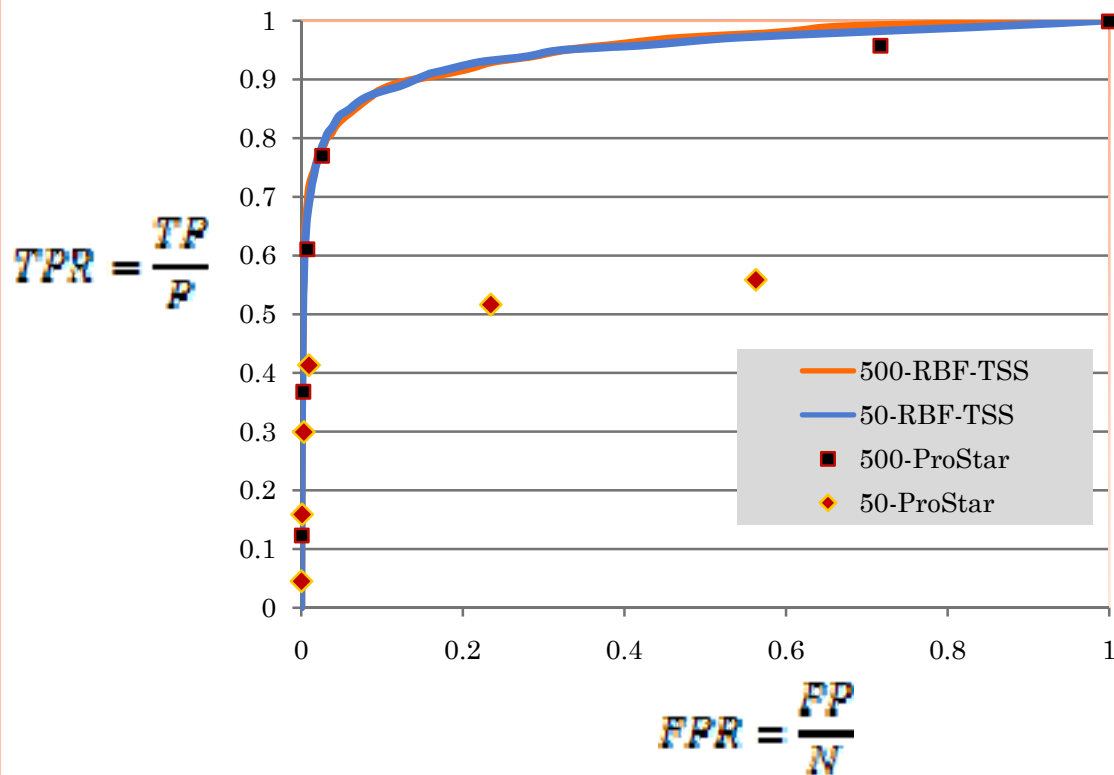
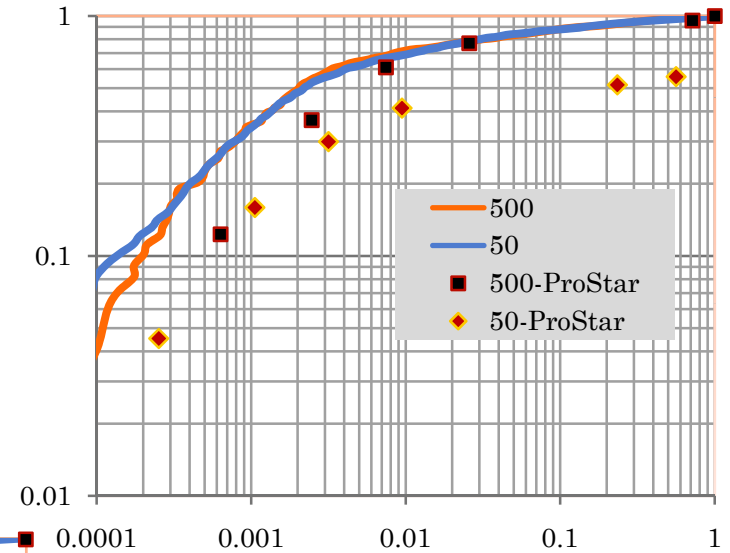
RBF-TSS VS. OTHER TOOLS

Curve	auROC %		auPRC %	
Chunk Size	50	500	50	500
RBF-TSS	94.75	95.08	24.08	54.64
ARTS	92.77	93.44	26.18	57.19
Eponine	88.48	91.51	11.79	40.80
McPromoter	92.55	93.59	6.32	24.23
FirstEF	71.29	90.25	6.54	40.89

	T	F
50	1588	1,087,666
500	943	108,782

- auPRC is found to be sensitive to having few negative samples with high scores.
- A removal of the 100 highest scoring negative samples resulted changes to the auPRC of 24.08% to 30.44% and 54.64% to 65.5% .
- While auROC changes were 94.75% to 94.76% and 95.08% to 95.14%

ROC (RBF-TSS vs. ProStar)



PRC (RBF-TSS vs. PROSTAR)

