ProSOM: human core promoter prediction

Thomas Abeel, Yvan Saeys and Yves Van de Peer

Department of Plant Systems Biology, VIB, Ghent University, Technologiepark 927, 9052 Gent, Belgium

Contact: thomas.abeel@psb.ugent.be
web: http://bioinformatics.psb.ugent.be

Short abstract

This poster presents ProSOM, a novel core promoter prediction program for the human genome with state-of-the-art performance. Self-organizing maps are used to distinguish promoter sequences from other sequences.
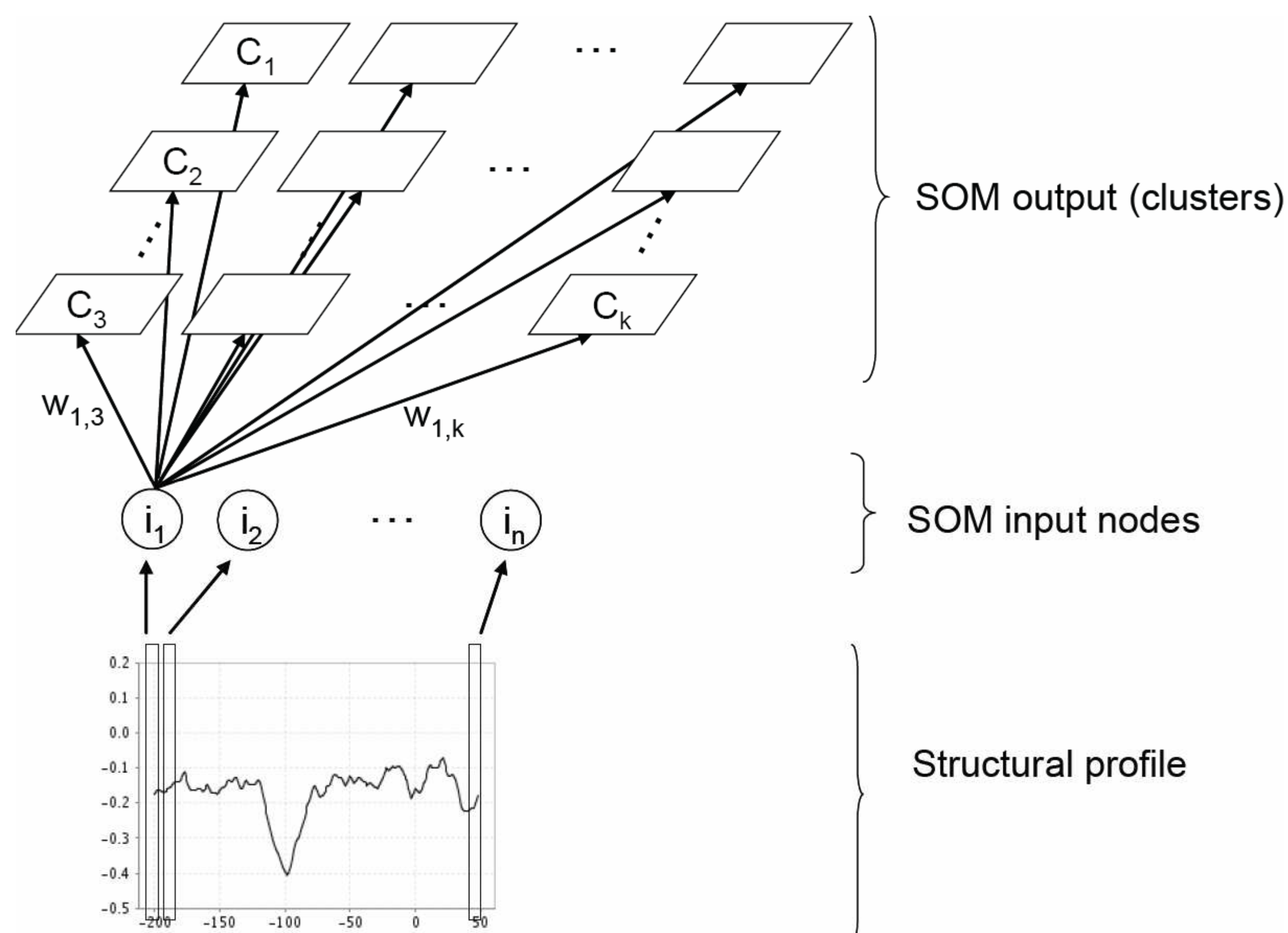
Core promoter prediction can provide guidance for experimental work or provide additional information for genome annotation projects.

Challenges

- Genome (3Gbp) → genes → promoter → core promoter (50 bp)
- Many genomes (350 projects)
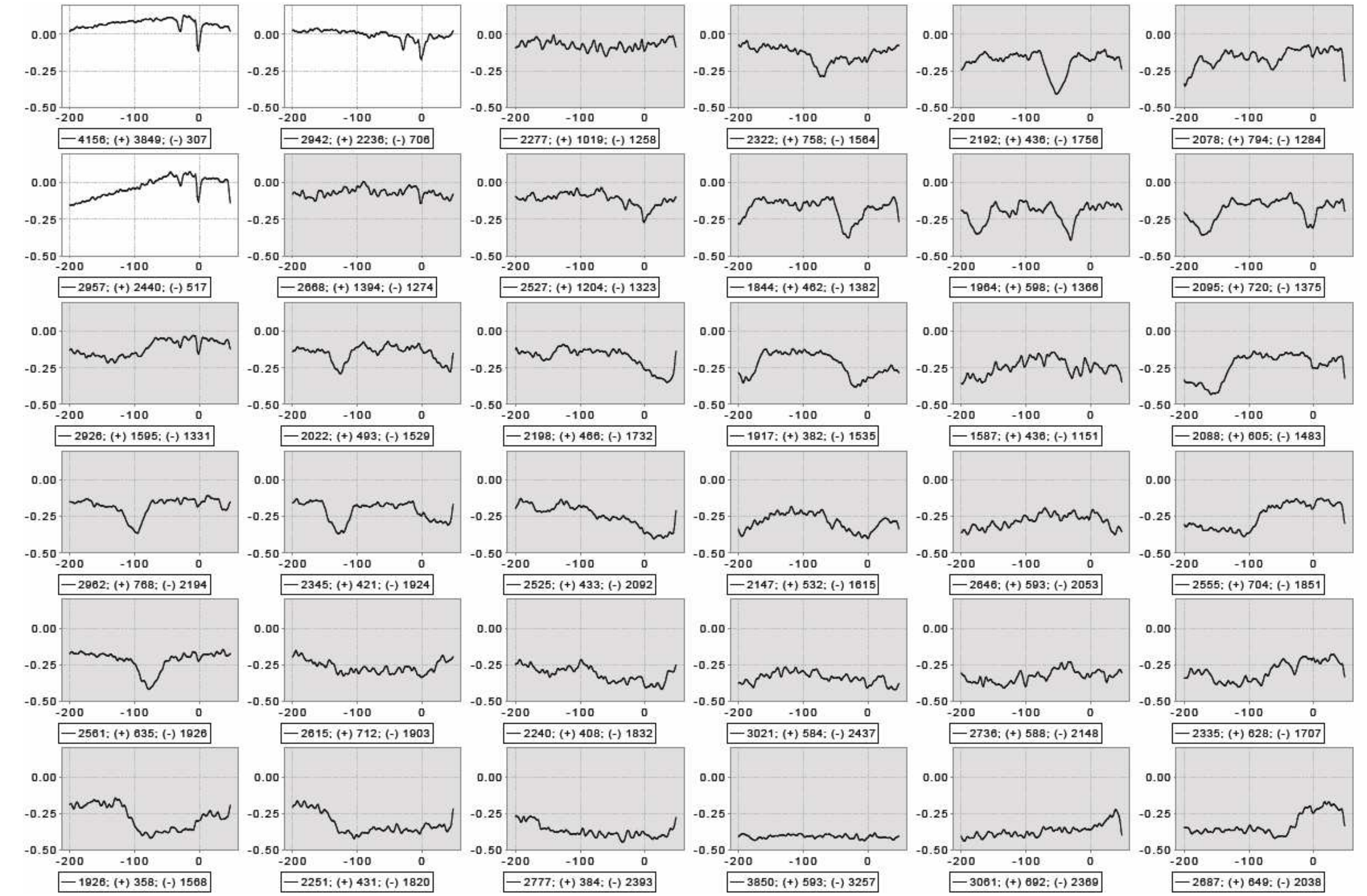- Little sequence preservation in core promoter
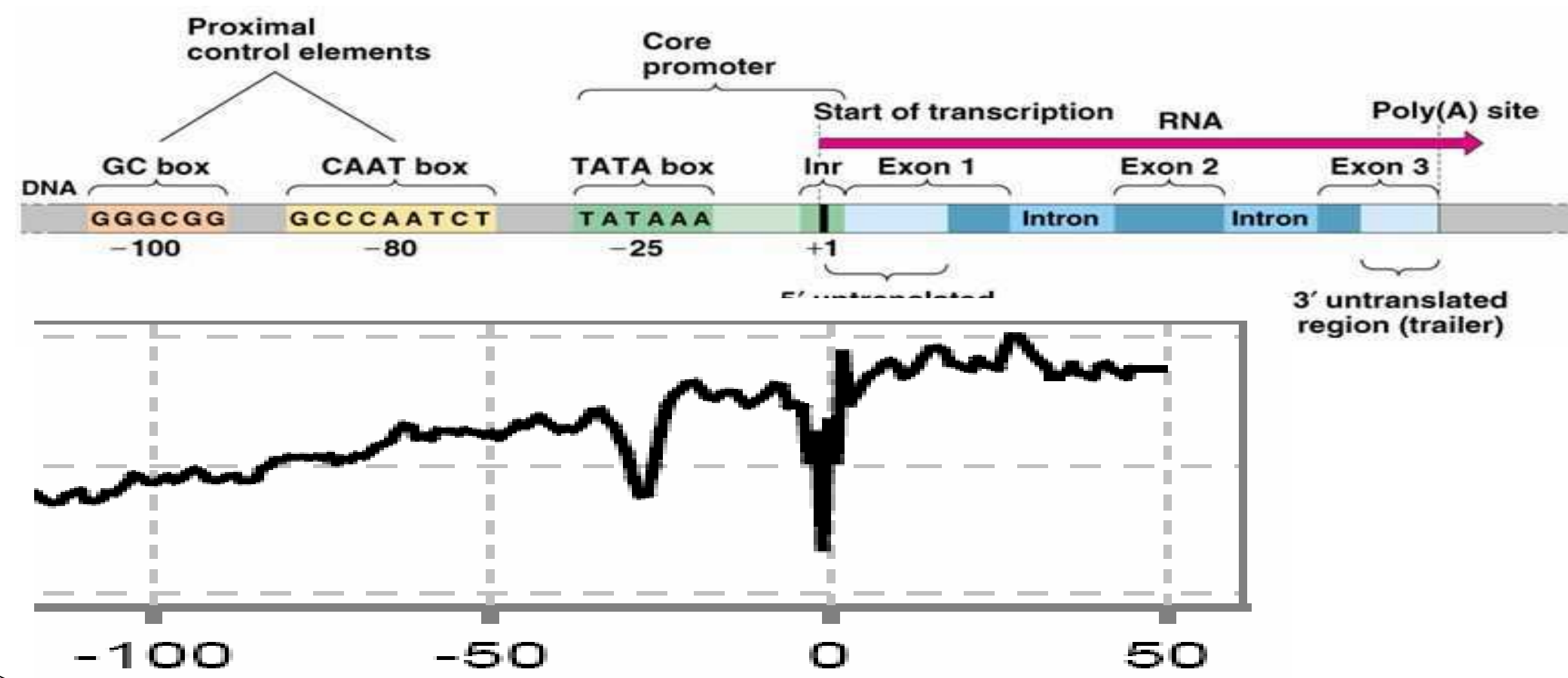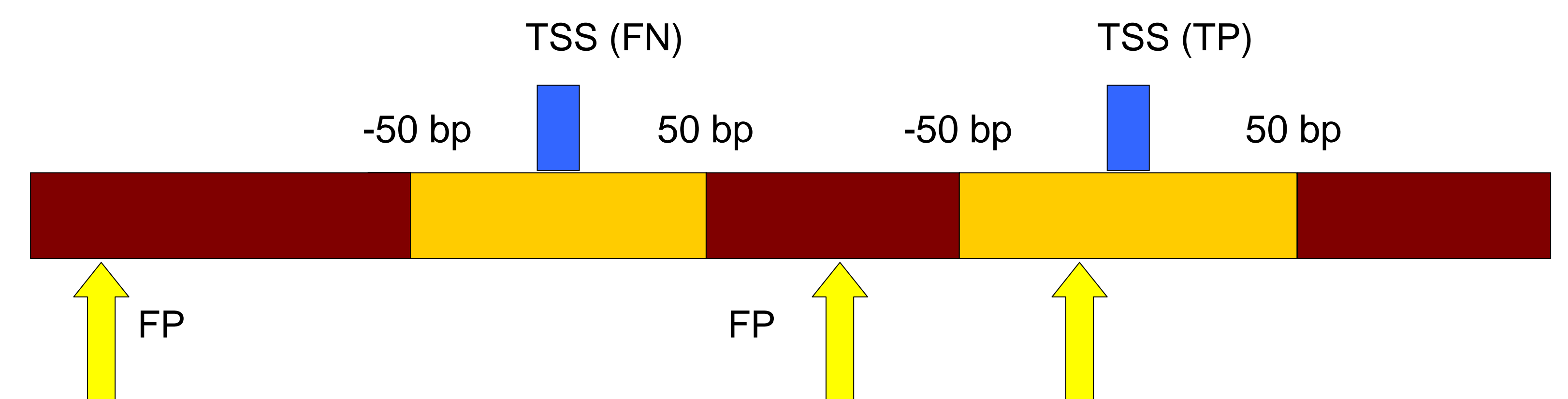

Self-organizing map

Clustering

- 30,000 promoters, 60,000 other sequences
- Clustered using a self-organizing map

- Three clusters with many promoters (white)
- When a sequence maps to these clusters we can predict a promoter

Validation

- Human genome assembly
- CAGE TSS tags for human (~120,000 TSS)
- TP: true positive: correctly predicted TSS
- FP: false positive: false prediction
- FN: false negative: unpredicted TSS

	50 bp distance	500 bp distance
recall	17%	38%
precision	30%	66%
F-measure	22%	48%

Abeel, T., **Saeys, Y.**, **Rouzé, P.**, **Van de Peer, Y.** (2008) ProSOM: Core promoter prediction based on unsupervised clustering of DNA physical profiles. Bioinformatics 24, i24-i31.