# Comparison of Support Vector Regression Models of

Transcription Factors E2F1 and E2F4's Binding Specificities to DNA Sequences

Sunwoo Yim[1] and Raluca Gordan[2]

[1]North Carolina School of Science and Mathematics, Durham, NC and [2]Duke Center for Genomic and Computational Biology

Introduction

Each cell plays a unique functional role in the body because of called transcription factors (TFs), or proteins that control how DNA information is interpreted.
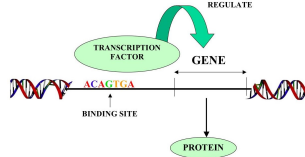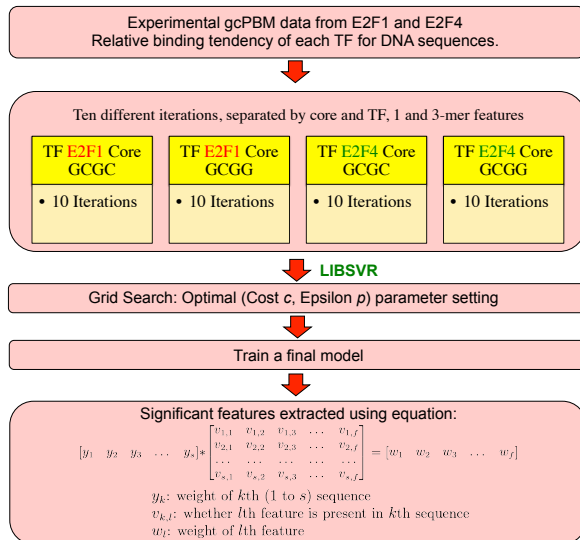

Source: Science 2.0. Three waves of innovation in vertebrate evolution, 2011.

TFs E2F1 and E2F4: Determine if normal healthy cells become benign or malignant tumors (cancerous).
- Share very similar structural domains
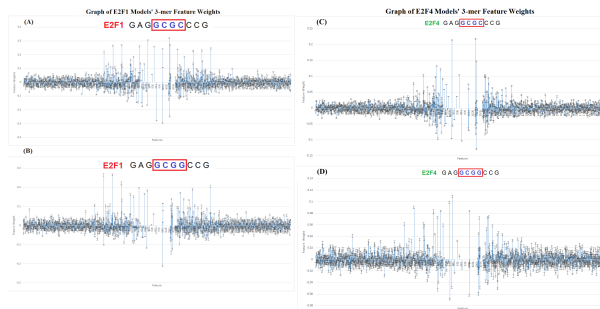- Bind to different DNA sequences- reason still unknown

Goal: Investigate the relationships between nucleotides in the TF binding sites to see how they affect E2F1 and E2F4 binding preferences.

Methods and Materials

Experimental gcPBM data from E2F1 and E2F4
Relative binding tendency of each TF for DNA sequences.

Ten different iterations, separated by core and TF, 1 and 3-mer features

TF E2F1 Core GCGC	TF E2F1 Core GCGG	TF E2F4 Core GCGC	TF E2F4 Core GCGG
• 10 Iterations	• 10 Iterations	• 10 Iterations	• 10 Iterations

LIBSVR

Grid Search: Optimal (Cost *c*, Epsilon *p*) parameter setting

Train a final model

Significant features extracted using equation:

$$[y_1 \ y_2 \ y_3 \ \dots \ y_s] * \begin{bmatrix} v_{1,1} & v_{1,2} & v_{1,3} & \dots & v_{1,f} \\ v_{2,1} & v_{2,2} & v_{2,3} & \dots & v_{2,f} \\ \dots & \dots & \dots & \dots & \dots \\ v_{s,1} & v_{s,2} & v_{s,3} & \dots & v_{s,f} \end{bmatrix} = [w_1 \ w_2 \ w_3 \ \dots \ w_f]$$

y_k: weight of kth (1 to s) sequence
$v_{k,l}$: whether lth feature is present in kth sequence
w_l: weight of lth feature

Results and Discussion

Graph of feature weights for the four categories

Most significant 3-mer features for the four categories

Significant 3-mer Features From Models
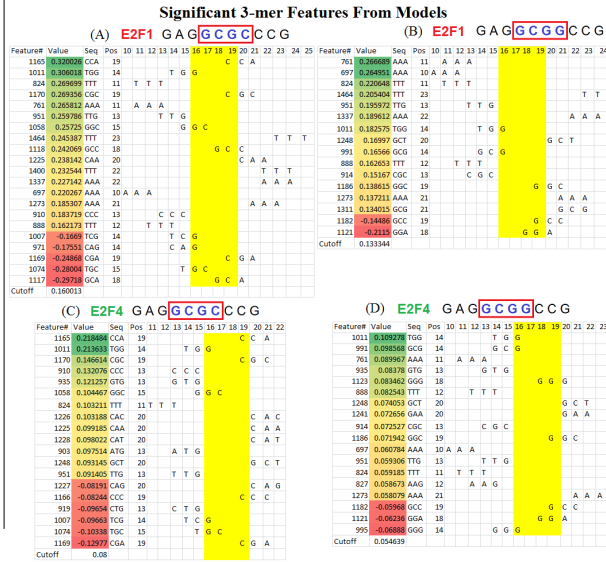
Comparison of the features between model

Graphs Comparing Feature Weights
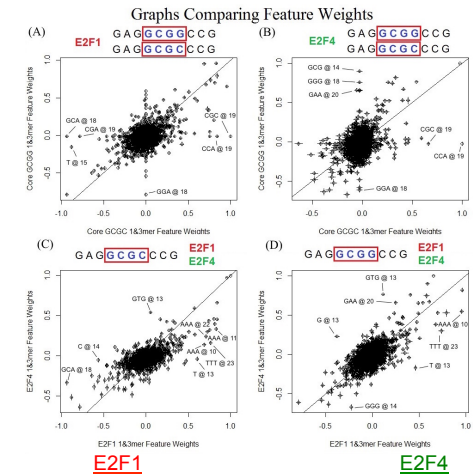
E2F1
- Greater affinity to sequences with A or T trinucleotides
- Tumors: increased latency and decreased frequency

E2F4
- Less predictable by models: external variables such as cofactors
- Doesn't induce apoptosis

Both
- 6-mer flanks have greatest effect on binding

Implications
- Determine binding affinities of E2F1 and E2F4 to DNA sites
- Predict likelihood of growing benign or malignant tumors

Conclusions

Contributions to scientific community:
- Differences in preferred features of TFs with similar structures
- Future research on cell proliferation and the development of cancerous tumors.

Future studies: Additional TFs in the E2F family to confirm results. Broaden scope to different families of TFs for more general conclusions.

Acknowledgements

For more information, contact: Sunwoo Yim, North Carolina School of Science and Mathematics, 1219 Broad Street, Durham, NC 27705

Email: yim16s@ncssm.edu   Web site: http://www.ncssm.edu