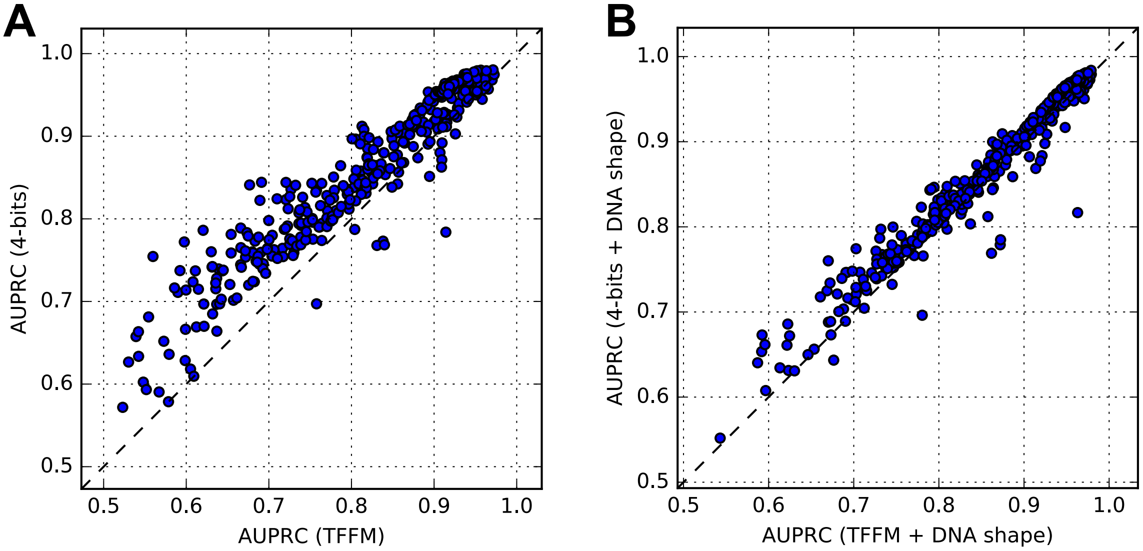
Data S5

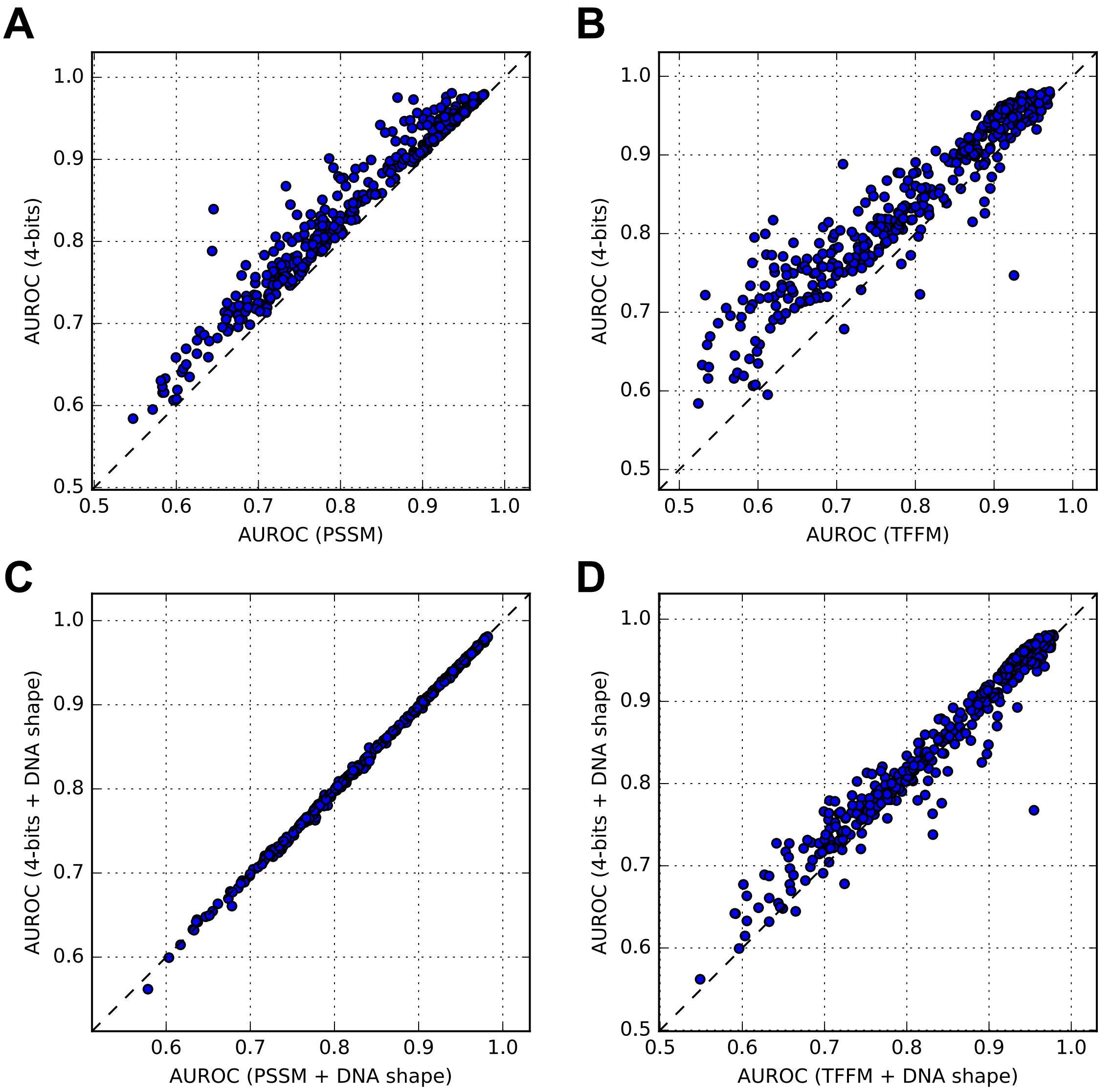
DNA shape features improve transcription factor binding site predictions *in vivo*

Anthony Mathelier[1,2], Beibei Xin[3], Tsu-Pei Chiu[3],
Lin Yang[3], Remo Rohs[3,*], and Wyeth W. Wasserman[1,*]

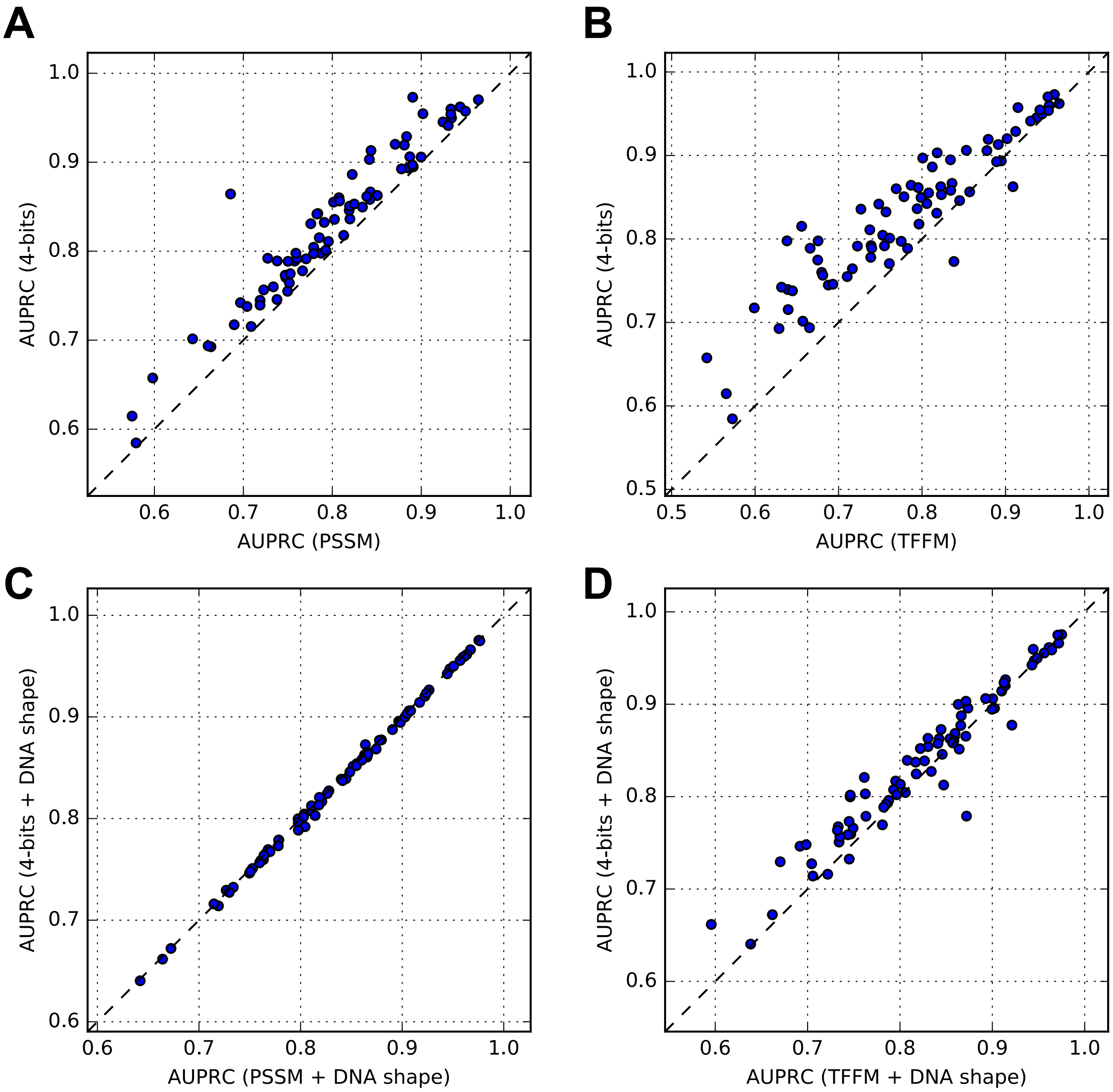
[1] Centre for Molecular Medicine at the Child and Family Research Institute,
Department of Medical Genetics, University of British Columbia,
980 West 28th Avenue, V5Z 4H4, Vancouver, BC, Canada
[2] Centre for Molecular Medicine Norway (NCMM), Nordic EMBL Partnership,
University of Oslo and Oslo University Hospital, Norway
[3] Molecular and Computational Biology Program, Departments of Biological Sciences,
Chemistry, Physics, and Computer Science,
University of Southern California, Los Angeles, CA 90089, USA
[*] Co-corresponding authors



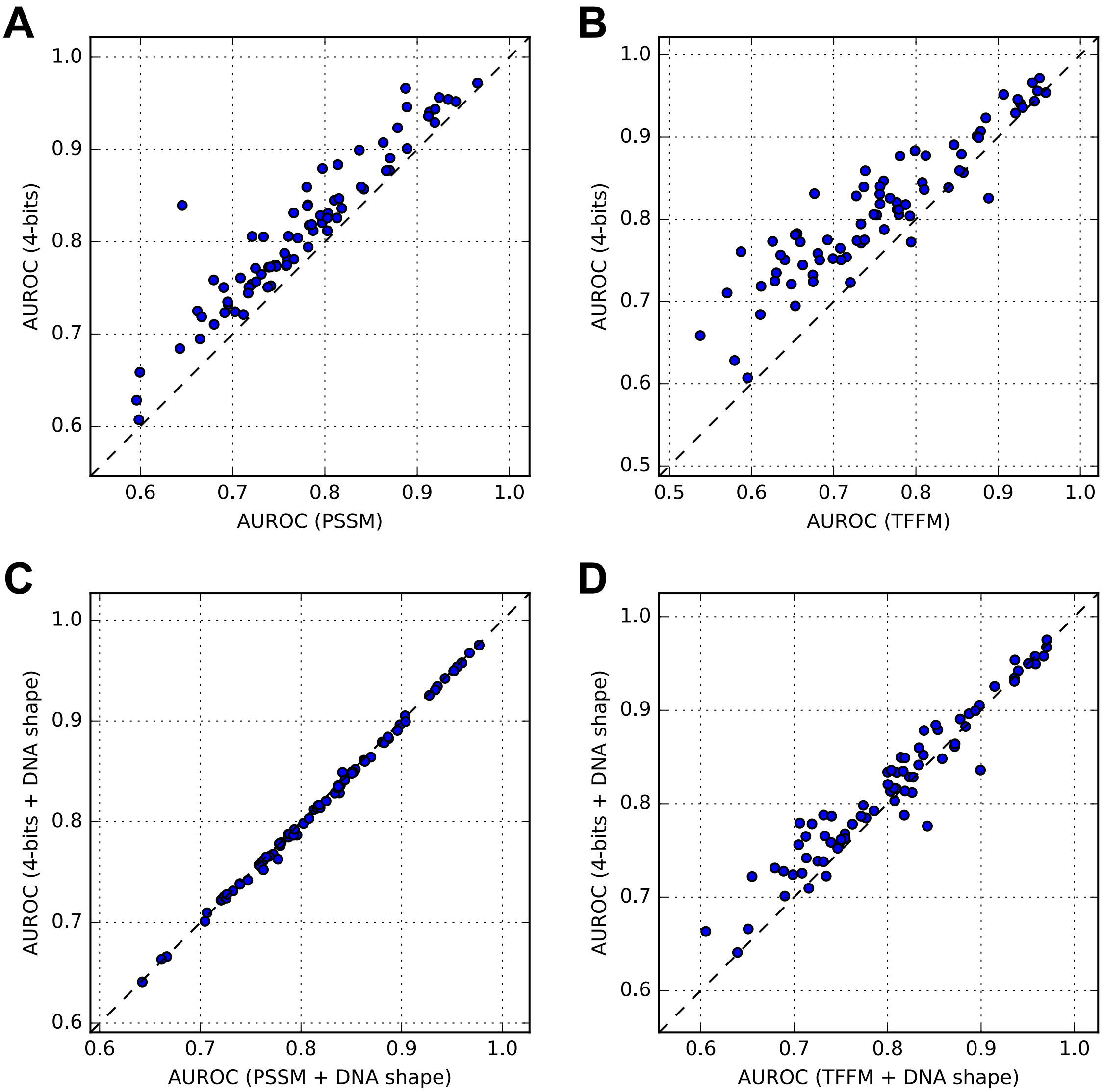
Related to Figure 3. Assessment of the predictive power of the TFFM and 4-bits approaches (see (Zhou et al.*, 2015) for the 4-bits encoding)to predict TFBSs in ChIP-seq regions. Comparison of the AUPRC obtained for the 400 human ENCODE ChIP-seq data sets when **A.** TFFM scores (x-axis) and 4-bits classifiers (y-axis) and **B.** TFFM + DNA shape (x-axis) and 4-bits + DNA shape (y-axis) classifiers are considered.*



Related to Figure 3. Assessment of the predictive power of the PSSM, TFFM, and 4-bits approaches (see (Zhou et al.*, 2015) for the 4-bits encoding)to predict TFBSs in ChIP-seq regions. Comparison of the AUROC obtained for the 400 human ENCODE ChIP-seq data sets when* ***A.*** *PSSM scores (x-axis) and 4-bits classifiers (y-axis),* ***B.*** *TFFM scores (x-axis) and 4-bits classifiers (y-axis),* ***C.*** *PSSM + DNA shape (x-axis) and 4-bits + DNA shape (y-axis) classifiers, and* ***D.*** *TFFM + DNA shape (x-axis) and 4-bits + DNA shape (y-axis) classifiers are considered.*



Related to Figure 3. Assessment of the predictive power of the PSSM, TFFM, and 4-bits approaches (see (Zhou et al.*, 2015) for the 4-bits encoding)to predict TFBSs in ChIP-seq regions. Comparison of the median AUPRC over all human ENCODE ChIP-seq data sets associated with each TF (one point per TF) when* ***A.*** *PSSM scores (x-axis) and 4-bits classifiers (y-axis),* ***B.*** *TFFM scores (x-axis) and 4-bits classifiers (y-axis),* ***C.*** *PSSM + DNA shape (x-axis) and 4-bits + DNA shape (y-axis) classifiers, and* ***D.*** *TFFM + DNA shape (x-axis) and 4-bits + DNA shape (y-axis) classifiers are considered.*



Related to Figure 3. Assessment of the predictive power of the PSSM, TFFM, and 4-bits approaches (see (Zhou et al., *2015) for the 4-bits encoding)to predict TFBSs in ChIP-seq regions. Comparison of the median AUPRC over all human ENCODE ChIP-seq data sets associated with each TF (one point per TF) when* ***A.*** *PSSM scores (x-axis) and 4-bits classifiers (y-axis),* ***B.*** *TFFM scores (x-axis) and 4-bits classifiers (y-axis),* ***C.*** *PSSM + DNA shape (x-axis) and 4-bits + DNA shape (y-axis) classifiers, and* ***D.*** *TFFM + DNA shape (x-axis) and 4-bits + DNA shape (y-axis) classifiers are considered.*