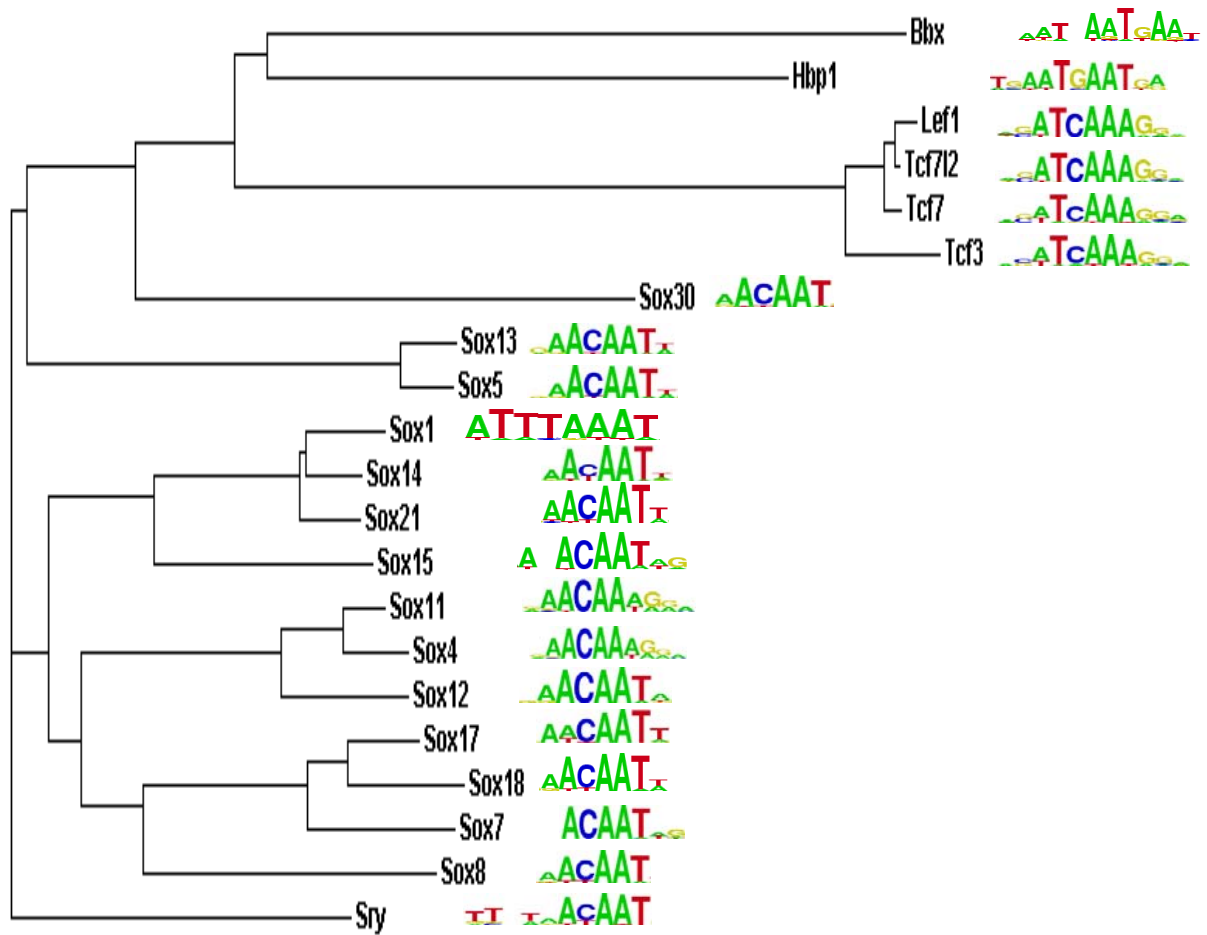


Figure S8. (A) HMG/SOX DNA-binding domains. *Top*, 2-D Hierarchical agglomerative clustering analysis of relative ranks for 310 8-mers x 21 HMG/SOX DNA-binding domains (with Sox7 as both DBD and FL). The 310 8-mers were selected because they have an E-score of 0.45 or greater for at least one of the DBDs shown. Each of the 310 8-mers was then given a rank score (between 1 and 310) within each column, and the ranks were analyzed here, in order to compensate for any overall differences in magnitude of the E-scores. *Bottom*, 6-mer sequences that are preferred within the 8-mers shown in the top panel. *Next page*, Seed-and-Wobble logos are shown next to a ClustalW phylogram derived using the amino-acid sequences of the DNA-binding domains.



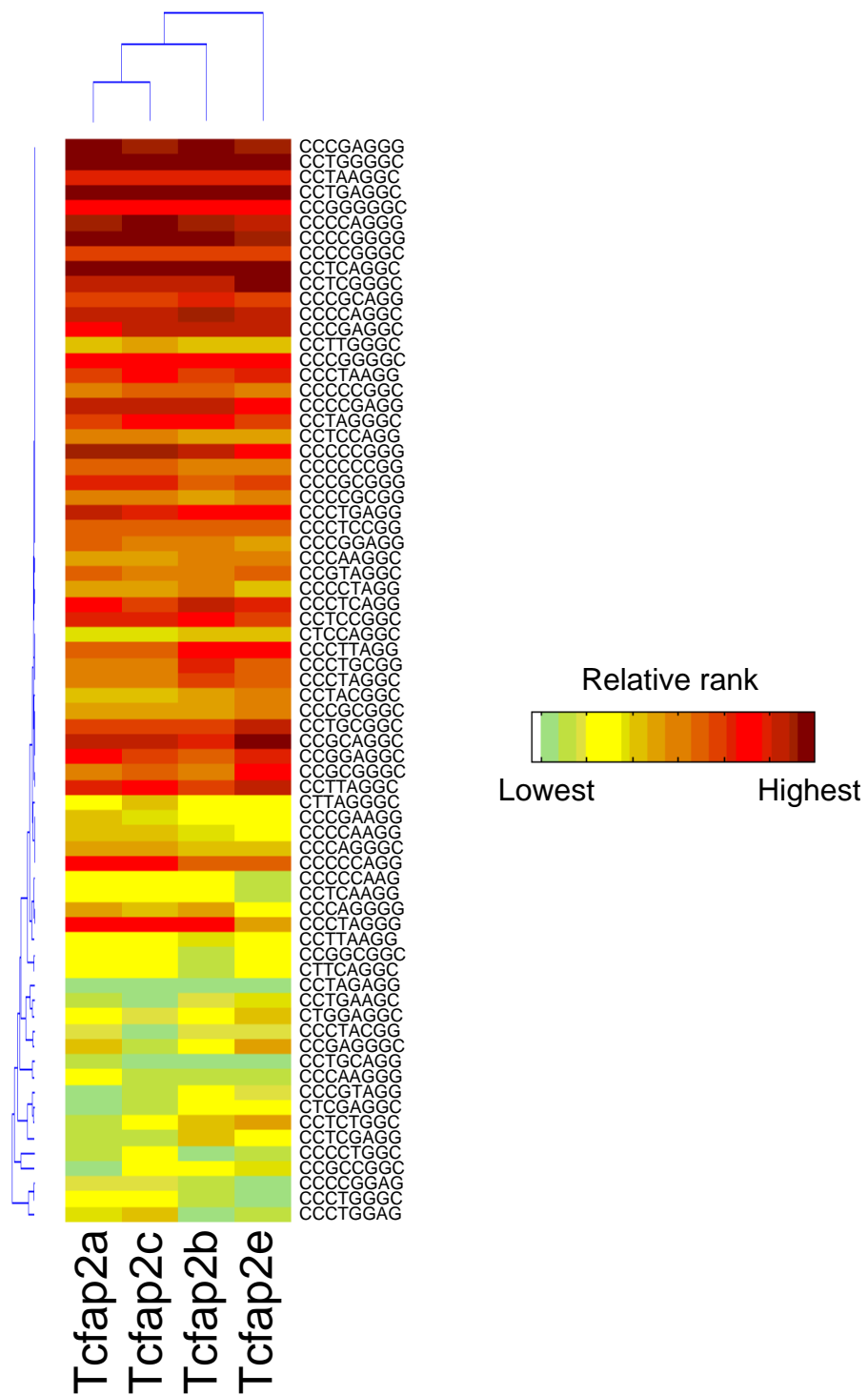


Figure S8. (B) AP-2 DNA-binding domains. 2-D Hierarchical agglomerative clustering analysis of relative ranks for 71 8-mers x 4 AP-2 DNA-binding domains. The 71 8-mers were selected because they have an E-score of 0.45 or greater for at least one of the TFs shown. Each of the 71 8-mers was then given a rank score (between 1 and 71) within each column and the ranks were analyzed, in order to compensate for any overall differences in magnitude of the E-scores.

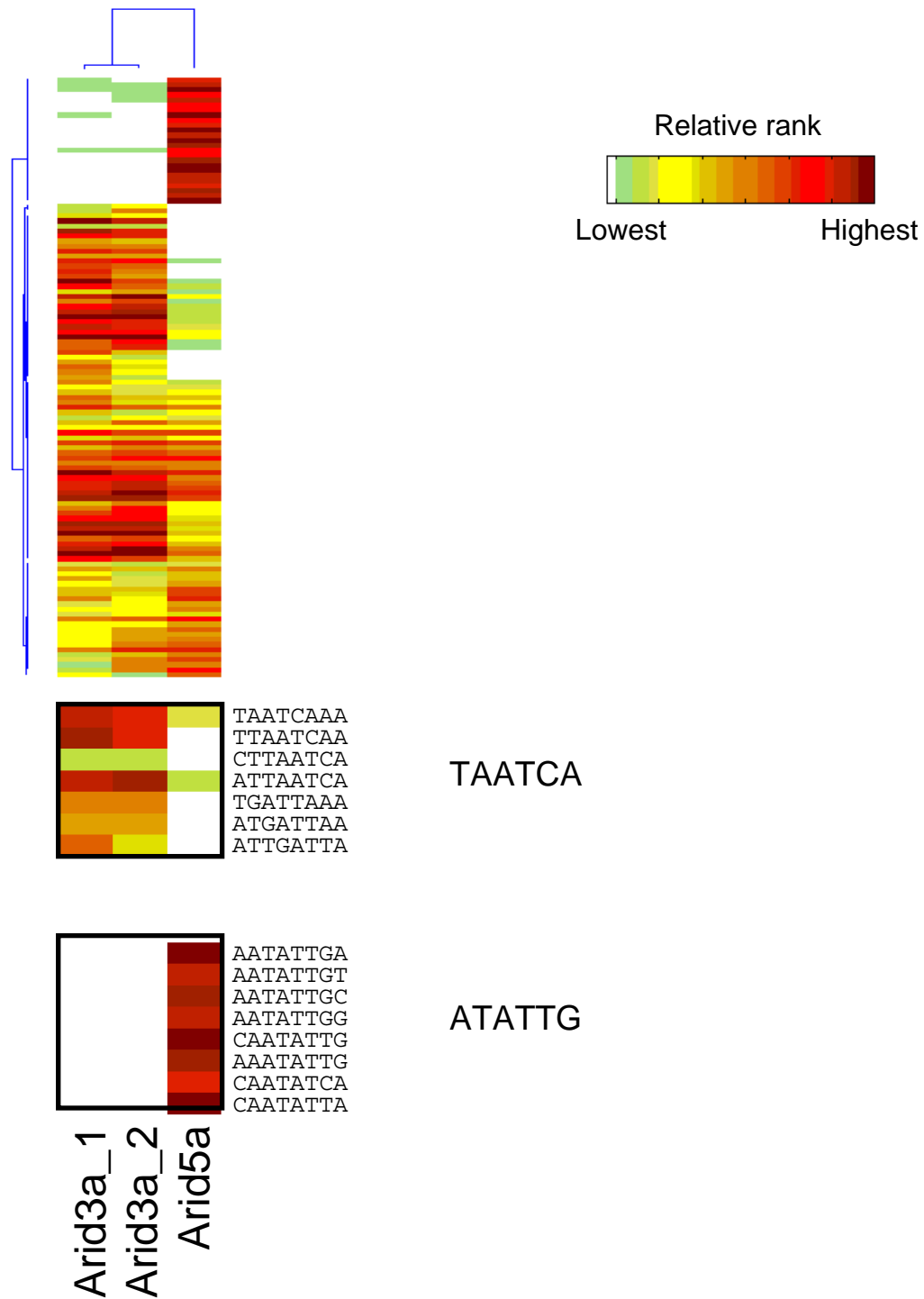


Figure S8. (C) ARID/BRIGHT DNA-binding domains. *Top*, 2-D Hierarchical agglomerative clustering analysis of relative ranks for 119 8-mers x 3 ARID/BRIGHT DNA-binding domains. The 119 8-mers were selected because they have an E-score of 0.45 or greater for at least one of the TFs shown. Each of the 119 8-mers was then given a rank score (between 1 and 119) within each column and the ranks were analyzed, in order to compensate for any overall differences in magnitude of the E-scores. *Bottom*, 6mer sequences that are preferred within the 8-mers shown in the top panel.

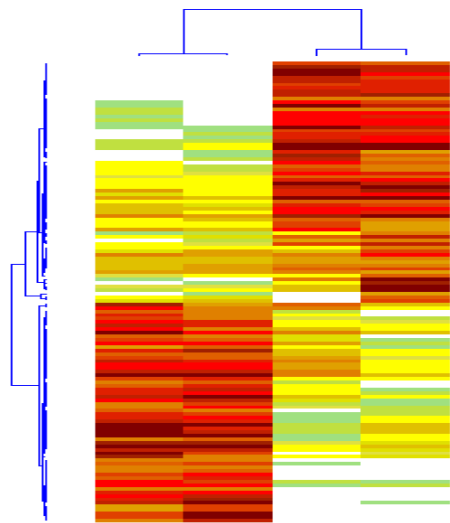
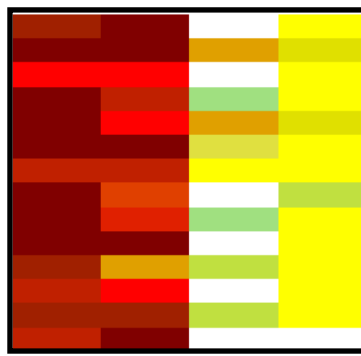


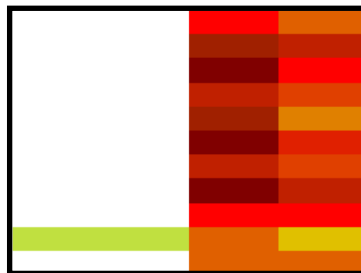
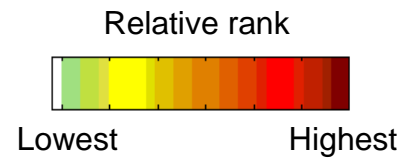
Figure S8. (D) BZIP DNA-binding domains. *Top*, 2-D Hierarchical agglomerative clustering analysis of relative ranks for 130 8-mers x 4 BZIP DNA-binding domains. The 130 8-mers were selected because they have an E-score of 0.45 or greater for at least one of the TFs shown. Each of the 130 8-mers was then given a rank score (between 1 and 130) within each column and the ranks were analyzed, in order to compensate for any overall differences in magnitude of the E-scores.

Middle, 6-mer sequences that are preferred within the 8-mers shown in the top panel. *Bottom*, Seed-and-Wobble logos are shown next to a ClustalW phylogram derived using the amino-acid sequences of the DNA-binding domains.



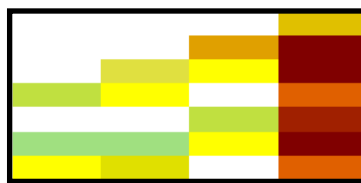
TCAGCAA
GTCAGCAA
ATCAGCAA
GTCAGCAC
CGTCAGCA
AGTCAGCA
AATCAGCA
GGTCAGCA
TGCTGACA
ATGCTGAC
CTGCTGAC
TATGCTGA
ATTGCTGA
AATGCTGA

TCAGCA



CGTCACCA
ACGTCACC
ACGTCACT
ACGTCACG
ACGTCACA
CACGTCAC
GACGTCAC
GTGACGTA
GTGACGCA
GTGACGAA
CGGTGACG

CGTCAC



GAGTCATC
GTGAGTCA
ATGAGTCA
CTGAGTCA
GTGACTCA
ATGACTCA
CTGACTCA

GAGTCA

Mafb
Mafk
Atf1
Jundm2

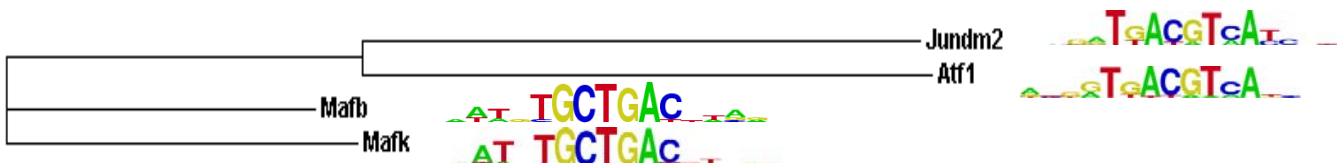
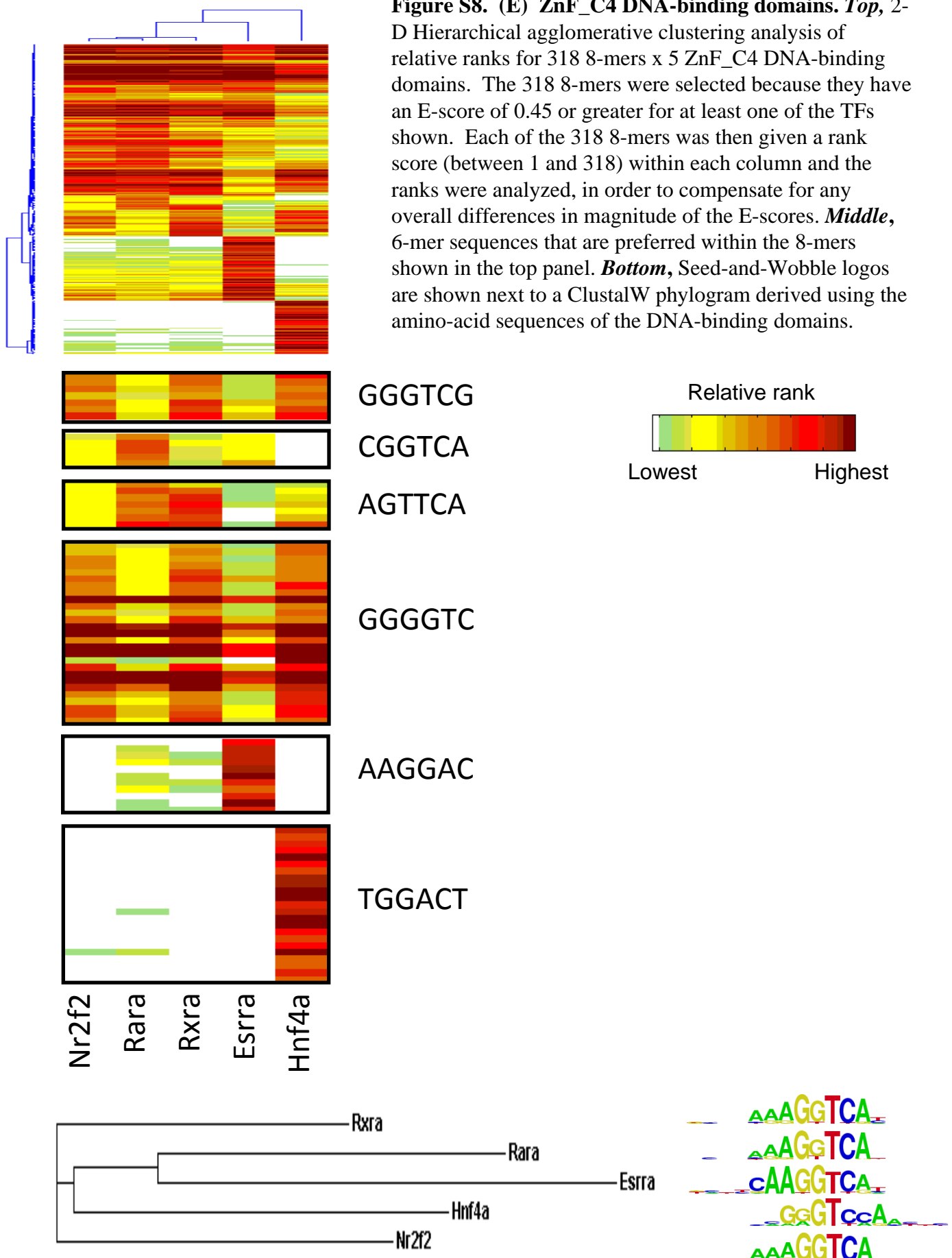


Figure S8. (E) ZnF_C4 DNA-binding domains. *Top*, 2-D Hierarchical agglomerative clustering analysis of relative ranks for 318 8-mers x 5 ZnF_C4 DNA-binding domains. The 318 8-mers were selected because they have an E-score of 0.45 or greater for at least one of the TFs shown. Each of the 318 8-mers was then given a rank score (between 1 and 318) within each column and the ranks were analyzed, in order to compensate for any overall differences in magnitude of the E-scores. *Middle*, 6-mer sequences that are preferred within the 8-mers shown in the top panel. *Bottom*, Seed-and-Wobble logos are shown next to a ClustalW phylogram derived using the amino-acid sequences of the DNA-binding domains.



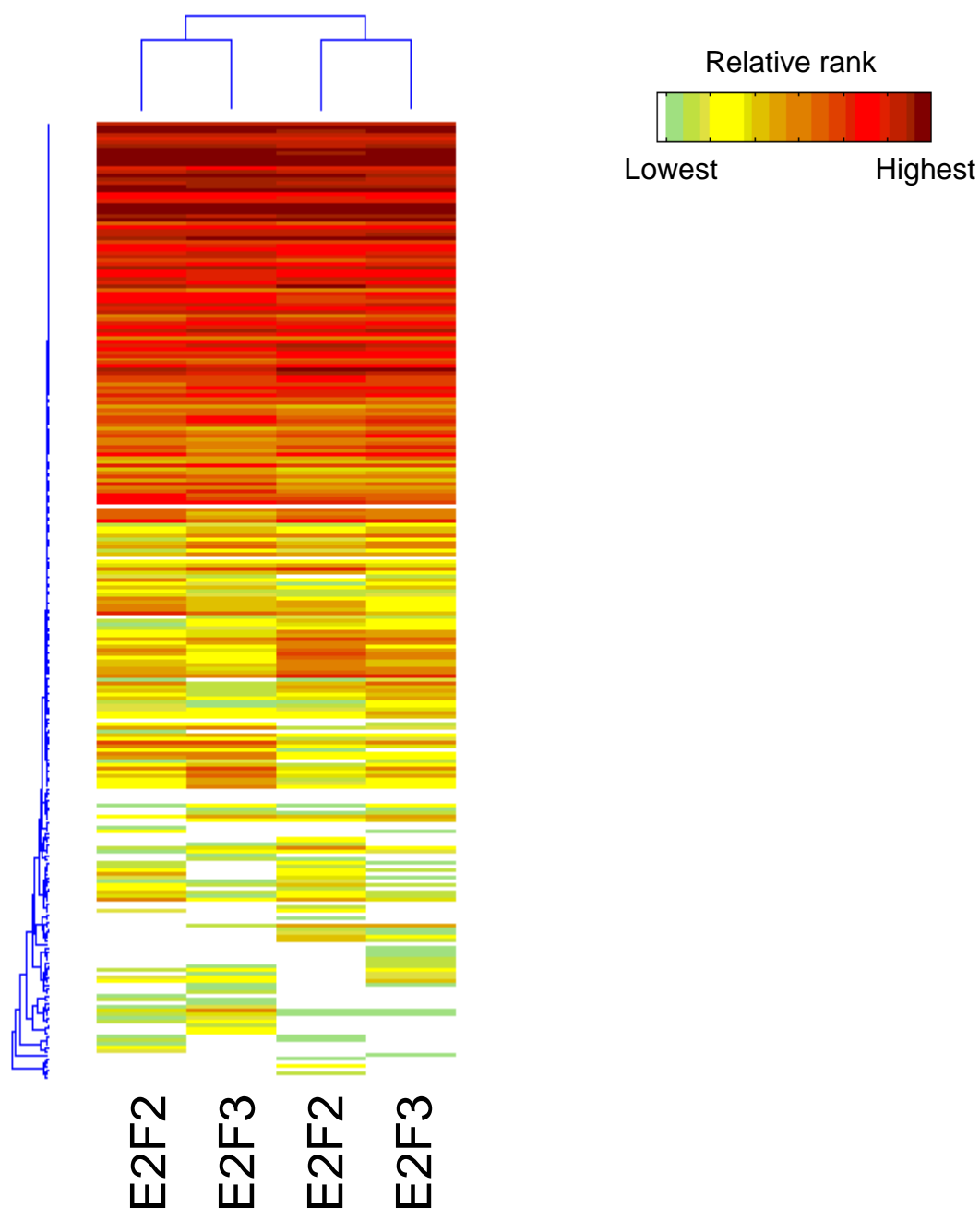
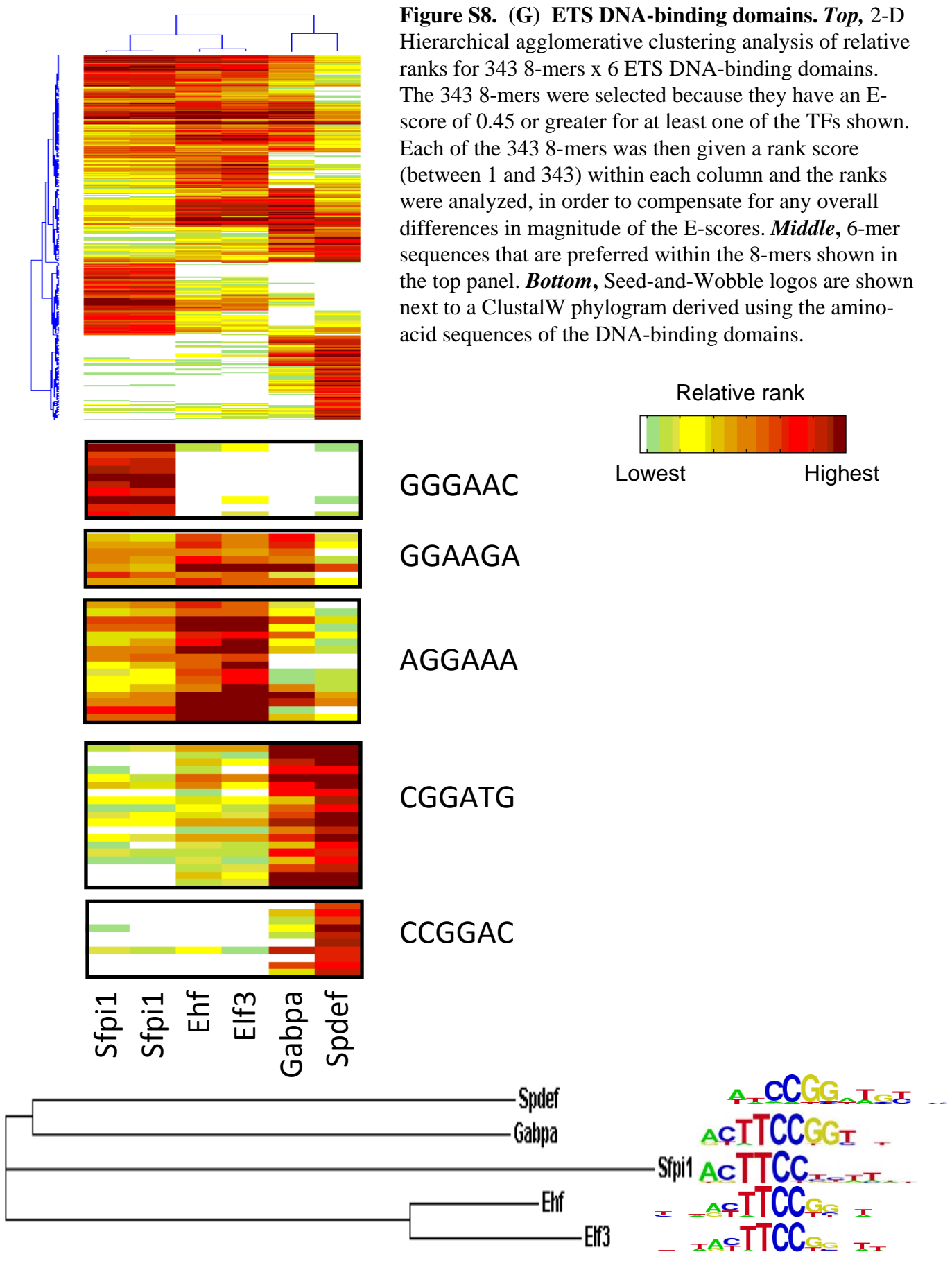


Figure S8. (F) E2F DNA-binding domains. 2-D Hierarchical agglomerative clustering analysis of relative ranks for 260 8-mers x 4 E2F DNA-binding domains. The 260 8-mers were selected because they have an E-score of 0.45 or greater for at least one of the TFs shown. Each of the 260 8-mers was then given a rank score (between 1 and 260) within each column and the ranks were analyzed, in order to compensate for any overall differences in magnitude of the E-scores.

Figure S8. (G) ETS DNA-binding domains. *Top*, 2-D Hierarchical agglomerative clustering analysis of relative ranks for 343 8-mers x 6 ETS DNA-binding domains. The 343 8-mers were selected because they have an E-score of 0.45 or greater for at least one of the TFs shown. Each of the 343 8-mers was then given a rank score (between 1 and 343) within each column and the ranks were analyzed, in order to compensate for any overall differences in magnitude of the E-scores. *Middle*, 6-mer sequences that are preferred within the 8-mers shown in the top panel. *Bottom*, Seed-and-Wobble logos are shown next to a ClustalW phylogram derived using the amino-acid sequences of the DNA-binding domains.



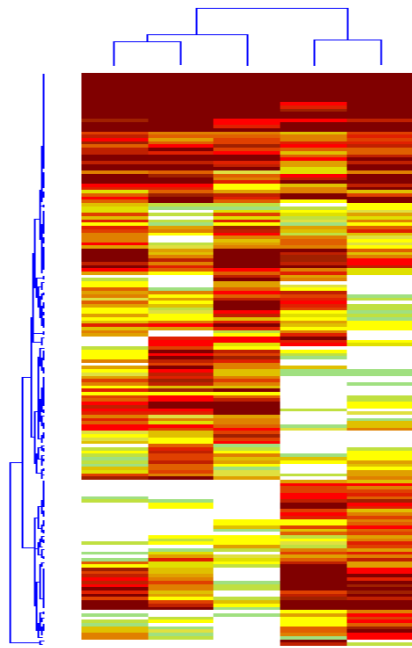
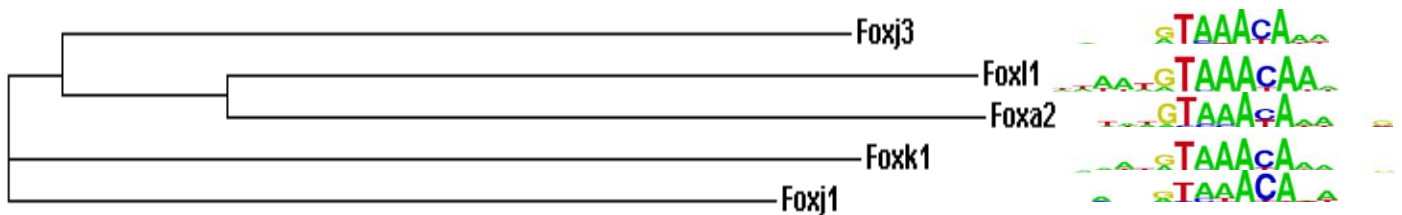
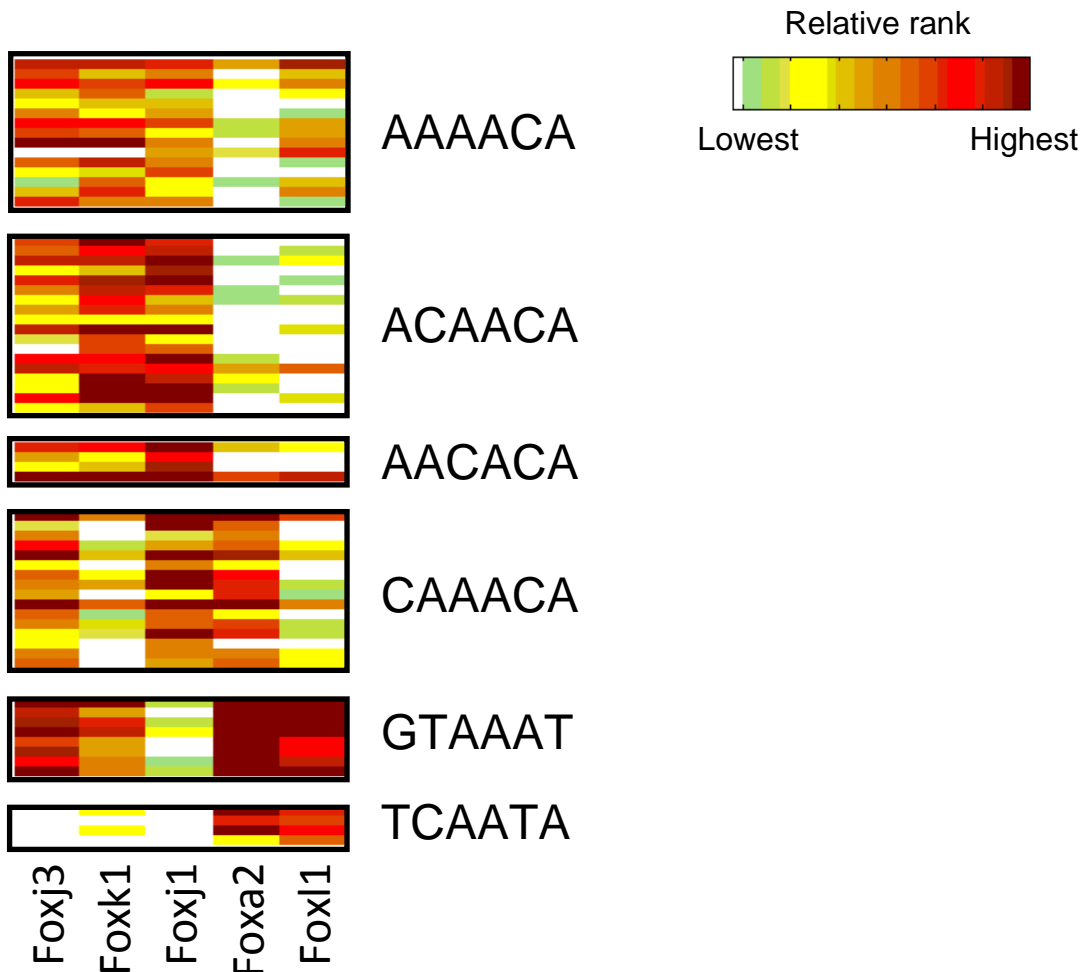


Figure S8. (H) Forkhead (FH) DNA-binding domains. *Top*, 2-D Hierarchical agglomerative clustering analysis of relative ranks for 176 8-mers x 5 FH DNA-binding domains. The 176 8-mers were selected because they have an E-score of 0.45 or greater for at least one of the TFs shown. Each of the 176 8-mers was then given a rank score (between 1 and 176) within each column and the ranks were analyzed, in order to compensate for any overall differences in magnitude of the E-scores. *Middle*, 6-mer sequences that are preferred within the 8mers shown in the top panel. *Bottom*, Seed-and-Wobble logos are shown next to a ClustalW phylogram derived using the amino-acid sequences of the DNA-binding domains.



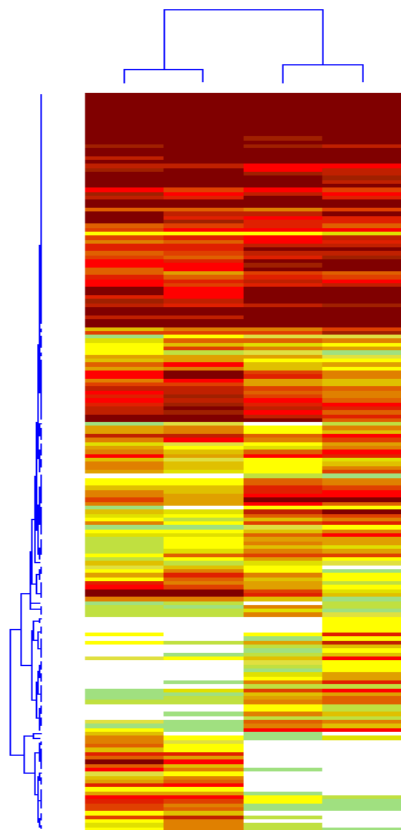
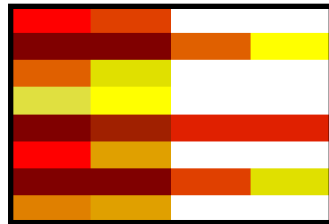
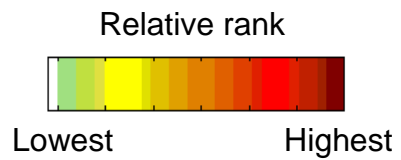


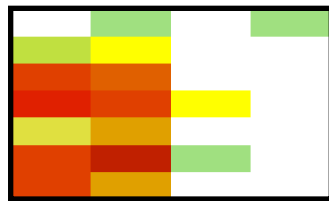
Figure S8. (I) GATA DNA-binding domains.

Top, 2-D Hierarchical agglomerative clustering analysis of relative ranks for 186 8-mers x 3 GATA DNA-binding domains (with Gata3 as both DBD and FL). The 186 8-mers were selected because they have an E-score of 0.45 or greater for at least one of the TFs shown. Each of the 186 8-mers was then given a rank score (between 1 and 186) within each column and the ranks were analyzed, in order to compensate for any overall differences in magnitude of the E-scores. **Middle**, 6-mer sequences that are preferred within the 8-mers shown in the top panel. **Bottom**, Seed-and-Wobble logos are shown next to a ClustalW phylogram derived using the amino-acid sequences of the DNA-binding domains.



ATCTGATC
 ATCTGATA
 AATCTGAT
 TAATCTGA
 TCAGATAA
 ATCAGATC
 ATCAGATA
 AATCAGAT

TCAGAT



AGATTAGC
 AGATTAAG
 AGATTAGA
 AGATTATC
 GAGATTAA
 ATAGATTA
 AGAGATTA

AGATTA

Gata3
 Gata3-FL
 Gata5
 Gata6



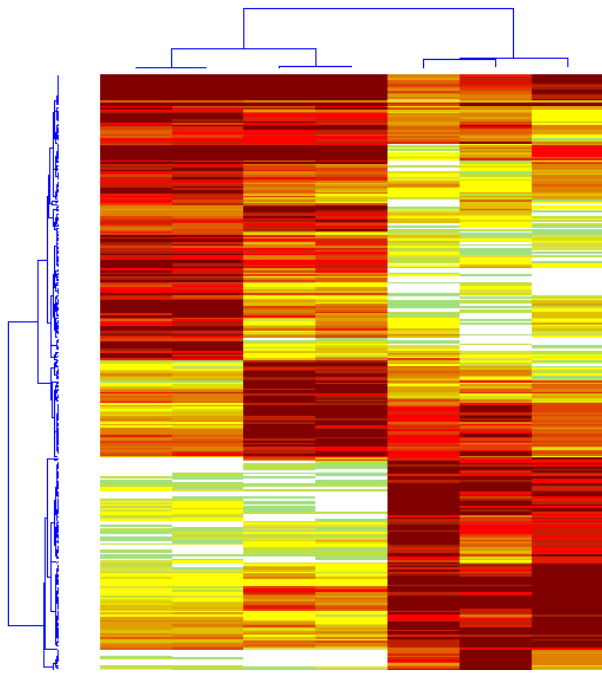


Figure S8. (J) HLH DNA-binding domains. *Top*, 2-D Hierarchical agglomerative clustering analysis of relative ranks for 320 8-mers x 6 HLH DNA-binding domains (with Max in duplicate and Bhlhb2 including DBD and FL). The 320 8-mers were selected because they have an E-score of 0.45 or greater for at least one of the TFs shown. Each of the 320 8-mers was then given a rank score (between 1 and 320) within each column and the ranks were analyzed, in order to compensate for any overall differences in magnitude of the E-scores. *Middle*, 6-mer sequences that are preferred within the 8-mers shown in the top panel. *Bottom*, Seed-and-Wobble logos are shown next to a ClustalW phylogram derived using the amino-acid sequences of the DNA-binding domains.

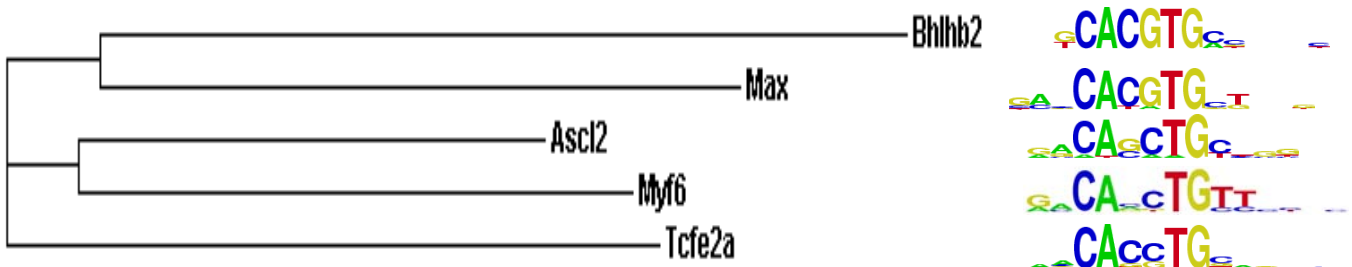
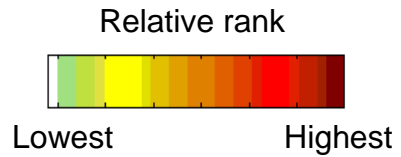
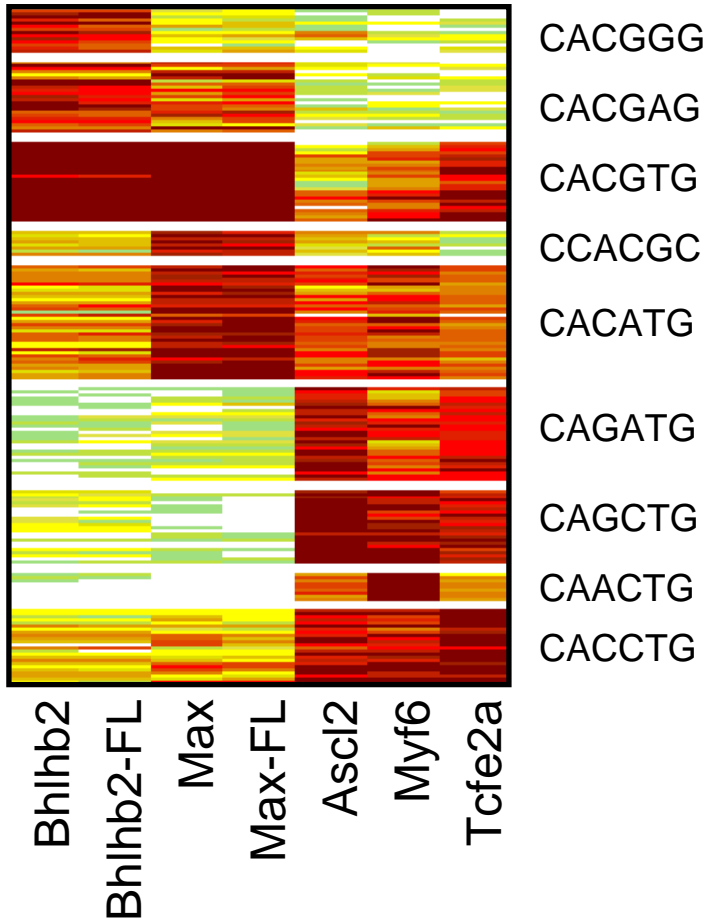
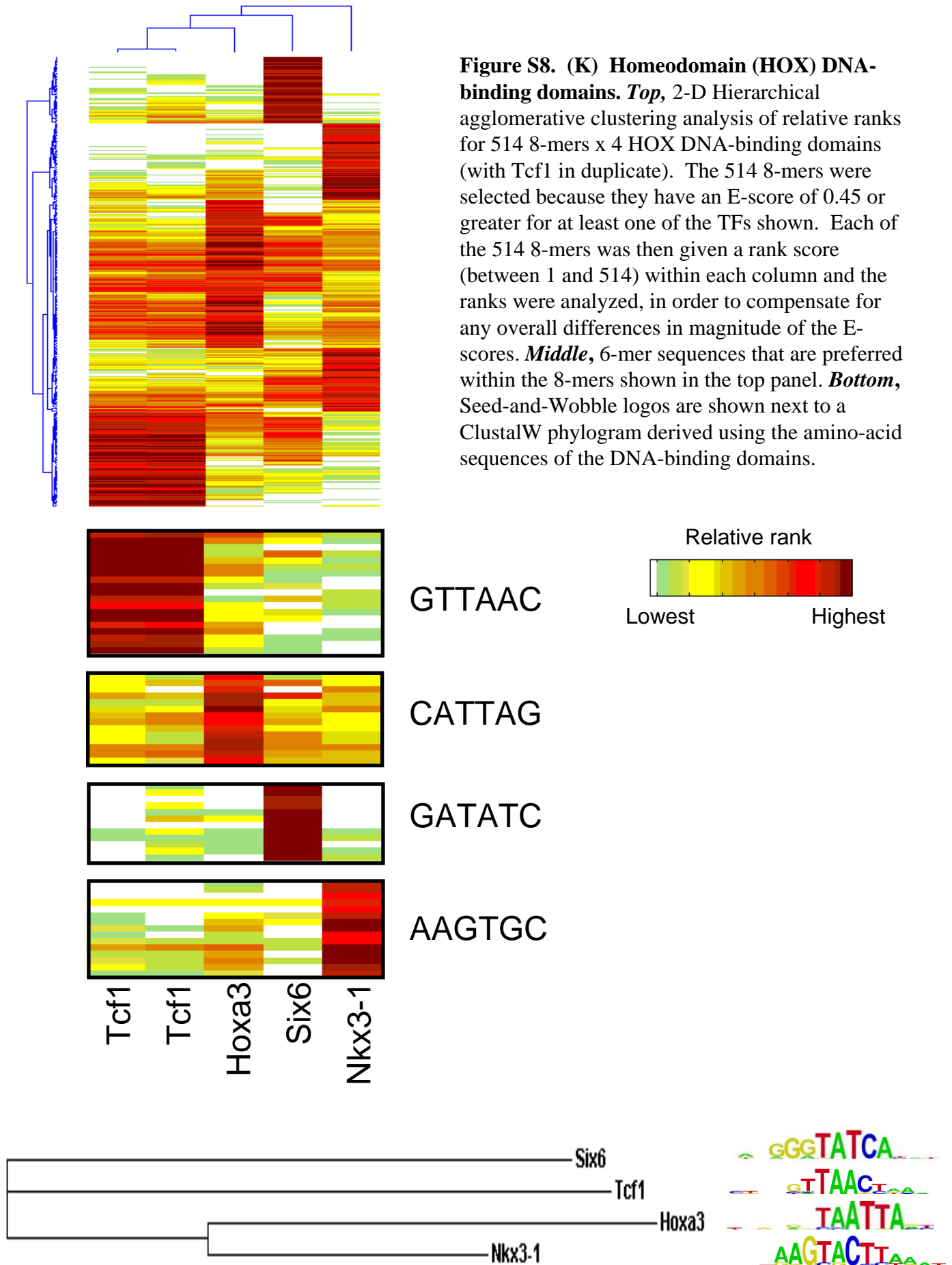


Figure S8. (K) Homeodomain (HOX) DNA-binding domains. *Top*, 2-D Hierarchical agglomerative clustering analysis of relative ranks for 514 8-mers x 4 HOX DNA-binding domains (with Tcf1 in duplicate). The 514 8-mers were selected because they have an E-score of 0.45 or greater for at least one of the TFs shown. Each of the 514 8-mers was then given a rank score (between 1 and 514) within each column and the ranks were analyzed, in order to compensate for any overall differences in magnitude of the E-scores. *Middle*, 6-mer sequences that are preferred within the 8-mers shown in the top panel. *Bottom*, Seed-and-Wobble logos are shown next to a ClustalW phylogram derived using the amino-acid sequences of the DNA-binding domains.



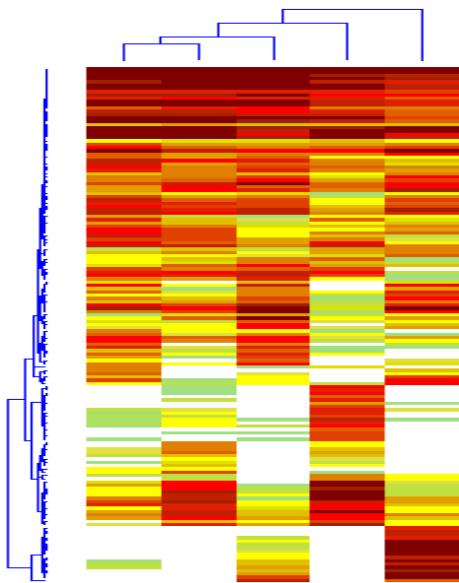
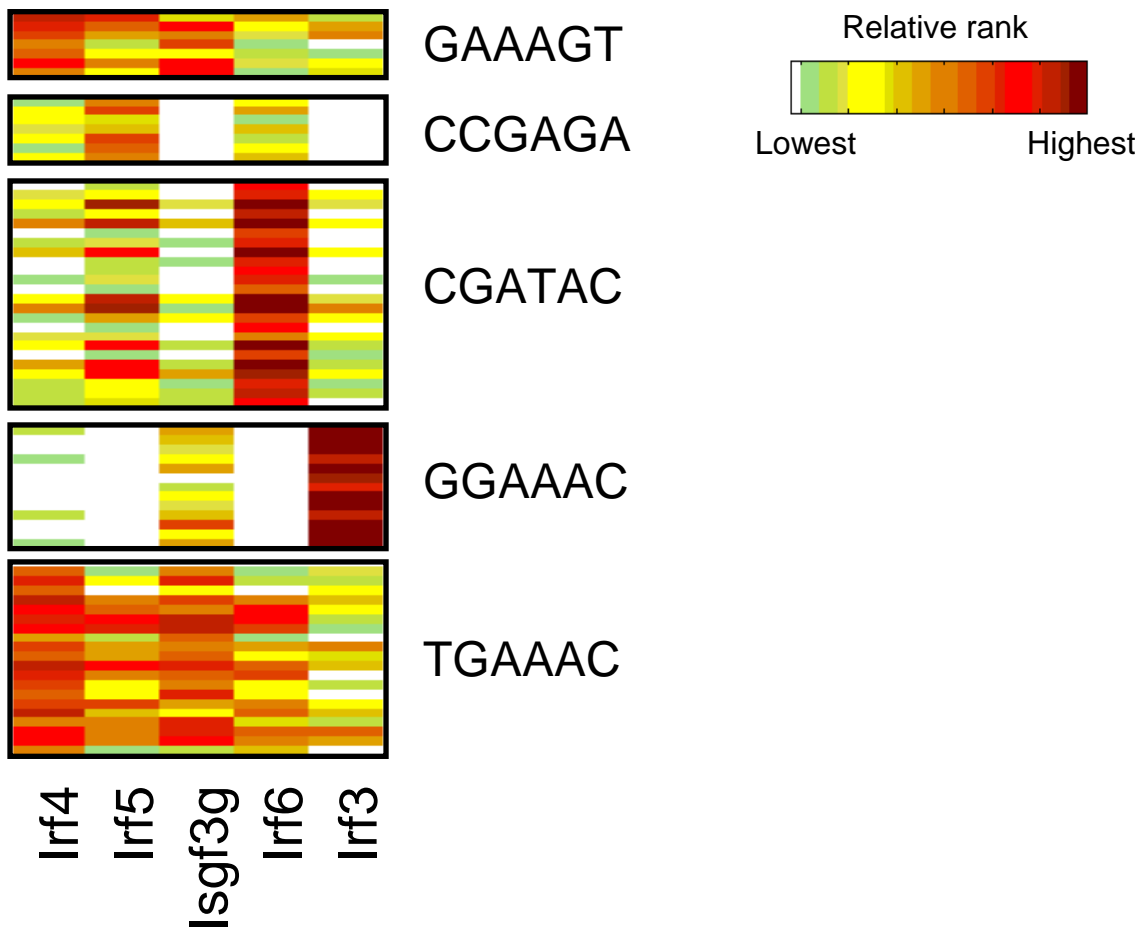


Figure S8. (L) IRF DNA-binding domains. *Top*, 2-D Hierarchical agglomerative clustering analysis of relative ranks for 157 8-mers x 5 IRF DNA-binding domains. The 157 8-mers were selected because they have an E-score of 0.45 or greater for at least one of the TFs shown. Each of the 157 8-mers was then given a rank score (between 1 and 157) within each column and the ranks were analyzed, in order to compensate for any overall differences in magnitude of the E-scores. *Middle*, 6-mer sequences that are preferred within the 8-mers shown in the top panel. *Bottom*, Seed-and-Wobble logos are shown next to a ClustalW phylogram derived using the amino-acid sequences of the DNA-binding domains.



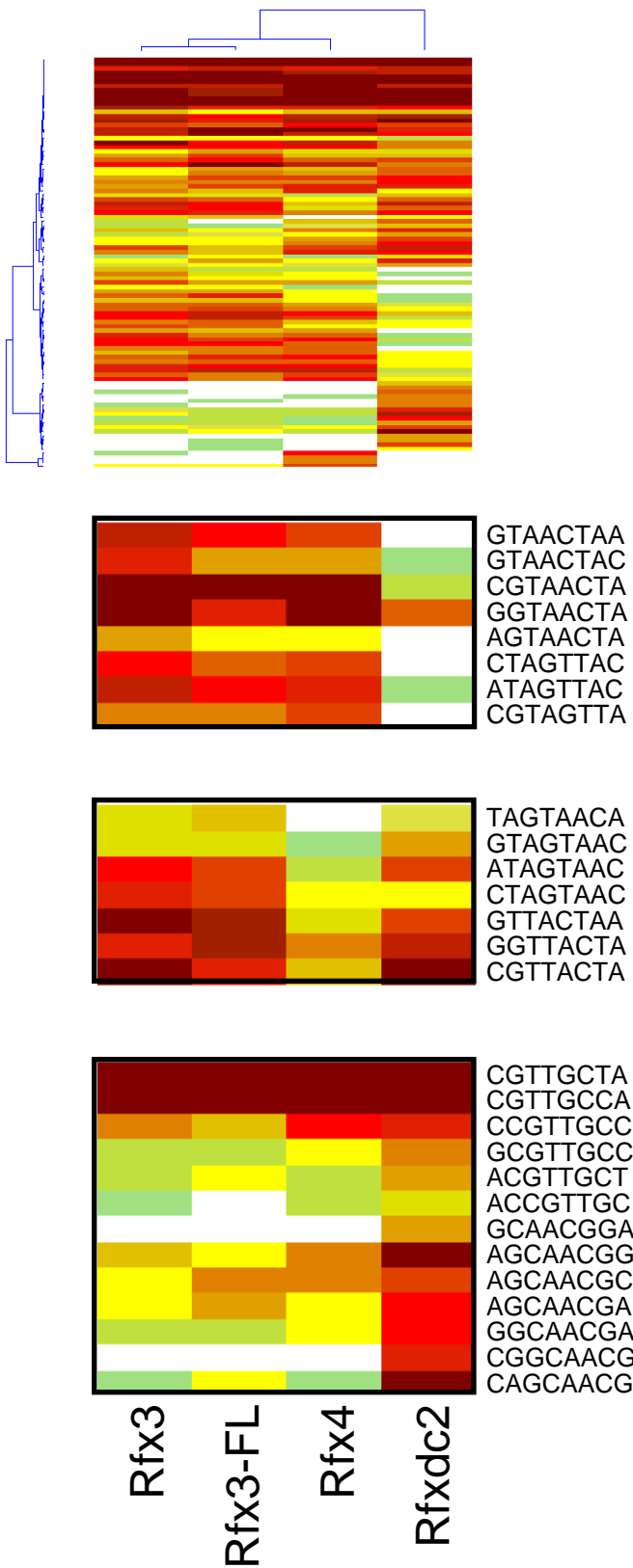
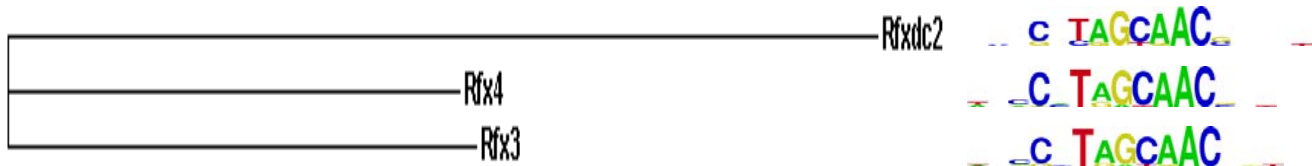


Figure S8. (M) RFX DNA-binding domains. *Top*, 2-D Hierarchical agglomerative clustering analysis of relative ranks for 94 8-mers x 3 IRF DNA-binding domains (with Rfx3 as both DBD and FL). The 94 8-mers were selected because they have an E-score of 0.45 or greater for at least one of the TFs shown. Each of the 94 8-mers was then given a rank score (between 1 and 94) within each column and the ranks were analyzed, in order to compensate for any overall differences in magnitude of the E-scores. ***Middle***, 6-mer sequences that are preferred within the 8-mers shown in the top panel. ***Bottom***, Seed-and-Wobble logos are shown next to a ClustalW phylogram derived using the amino-acid sequences of the DNA-binding domains.



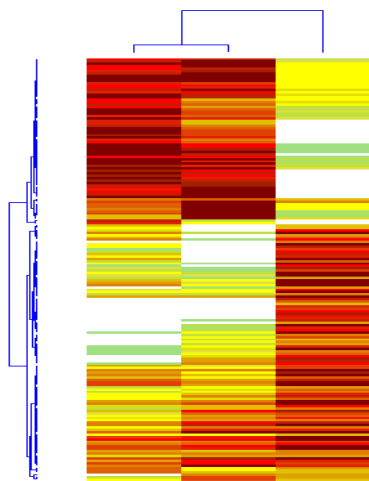
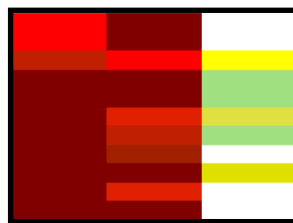
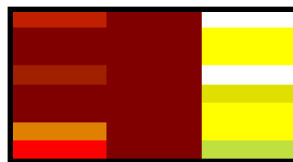


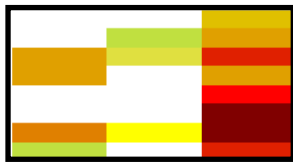
Figure S8. (N) SAND DNA-binding domains. *Top*, 2-D Hierarchical agglomerative clustering analysis of relative ranks for 178 8-mers x 3 SAND DNA-binding domains. The 178 8-mers were selected because they have an E-score of 0.45 or greater for at least one of the TFs shown. Each of the 178 8-mers was then given a rank score (between 1 and 178) within each column and the ranks were analyzed, in order to compensate for any overall differences in magnitude of the E-scores. ***Middle***, 6-mer sequences that are preferred within the 8-mers shown in the top panel. ***Bottom***, Seed-and-Wobble logos are shown next to a ClustalW phylogram derived using the amino-acid sequences of the DNA-binding domains.



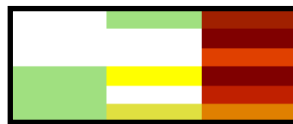
CGACAA



CGGAAA



ACGTAG



ACGCAC

RC900814
Sp100
Gmeb1

