SUPPLEMENTARY DATA

TFBSshape: a motif database for DNA shape features of transcription factor binding sites

Lin Yang[1], Tianyin Zhou[1], Iris Dror[1,2], Anthony Mathelier[3], Wyeth W. Wasserman[3], Raluca Gordân[4], and Remo Rohs[1,*]

[1] Molecular and Computational Biology Program, University of Southern California, Los Angeles, CA 90089, USA

[2] Department of Biology, Technion – Israel Institute of Technology, Technion City, Haifa 32000, Israel

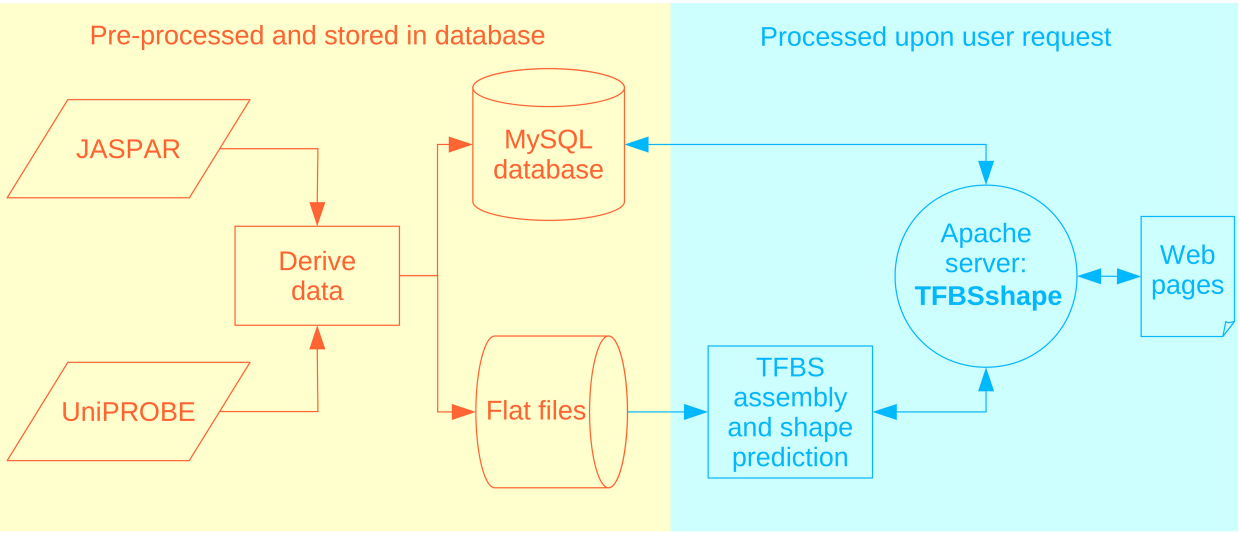
[3] Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada

[4] Institute for Genome Sciences & Policy, Duke University, Durham, NC 27708, USA

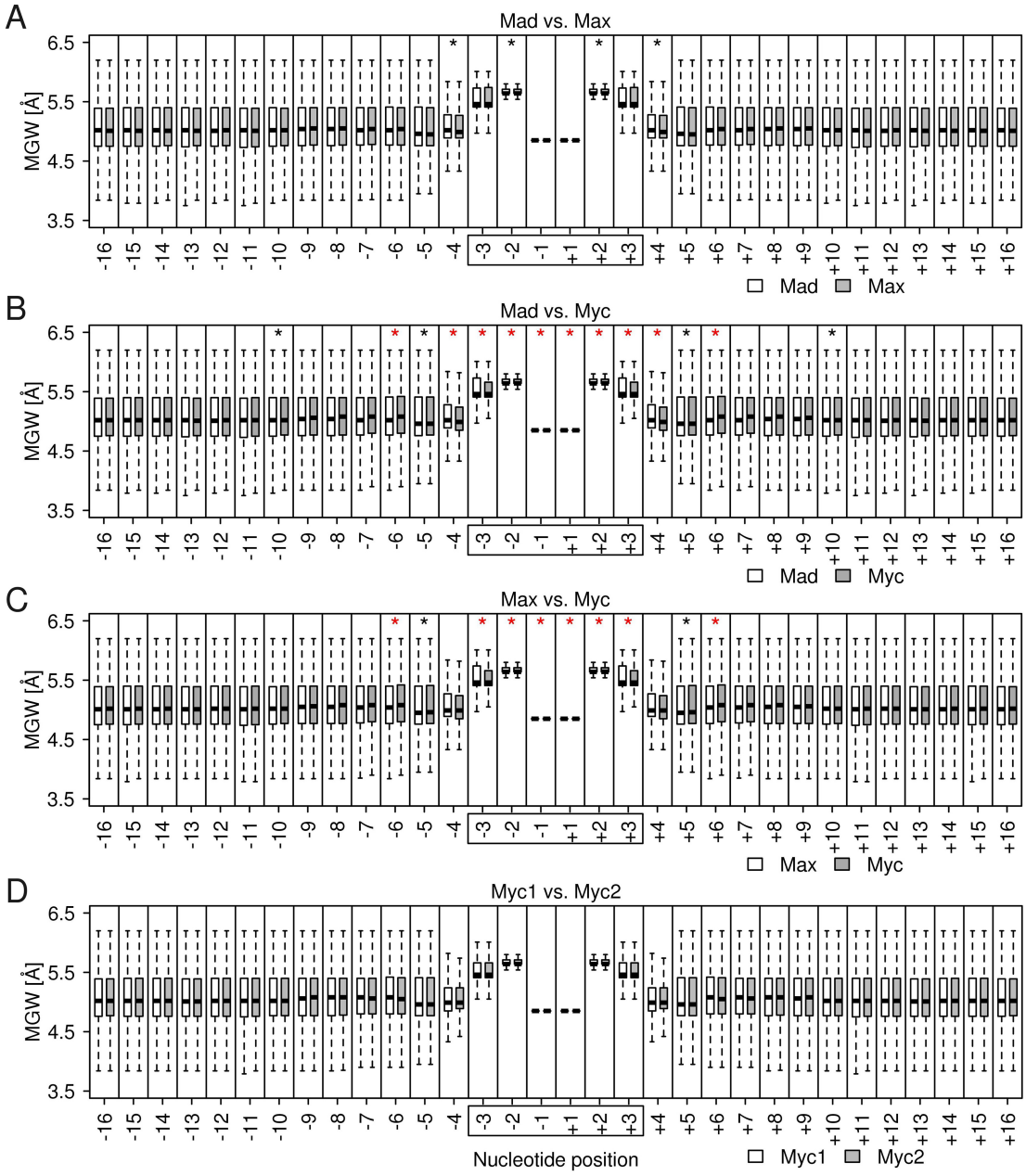
*To whom correspondence may be addressed:

Remo Rohs, Ph.D.
Molecular and Computational Biology Program
1050 Childs Way RRI 404C
University of Southern California
Los Angeles, CA 90089
United States
Tel: +1-213-740-0552
Fax: +1-213-821-4257
Email: rohs@usc.edu

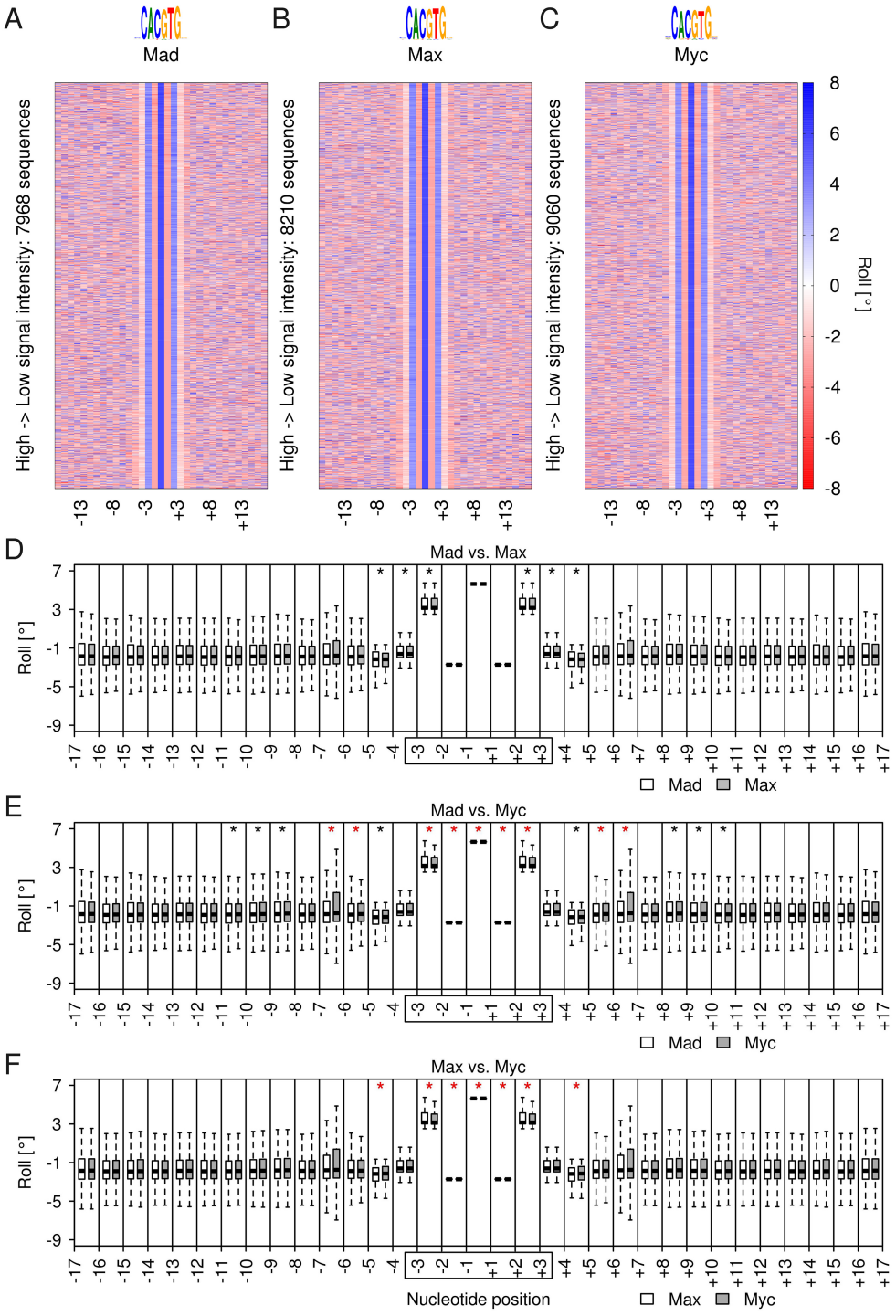
SUPPLEMENTARY FIGURES



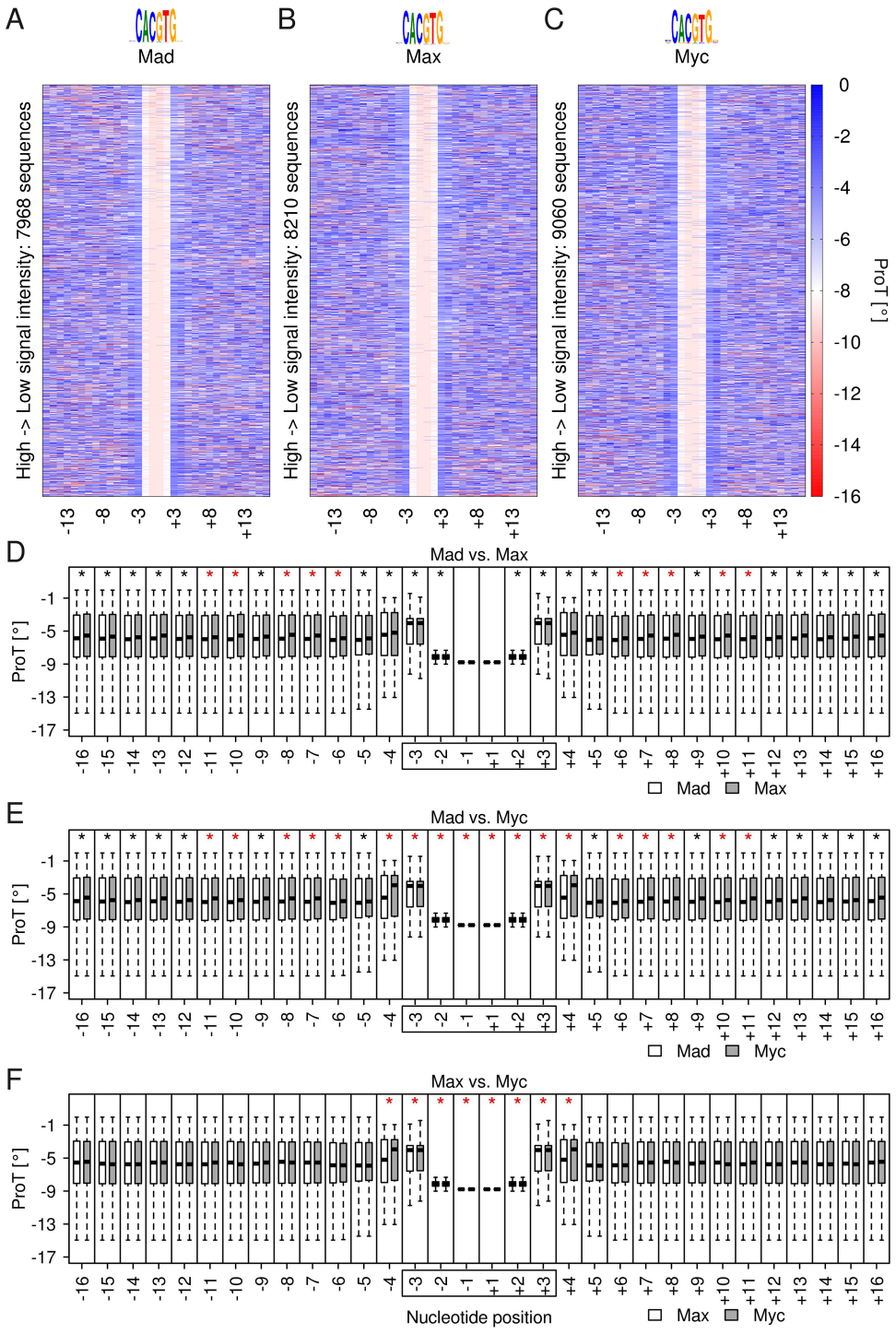
Supplementary Figure S1. TFBSshape database flowchart. Input data are nucleotide sequences derived from the motif databases JASPAR and UniPROBE, which are stored and managed using a MySQL database (yellow). Following this pre-processing, TFBSs are assembled, DNA shape features are predicted “on the fly”, and an Apache server provides the user interface (blue).



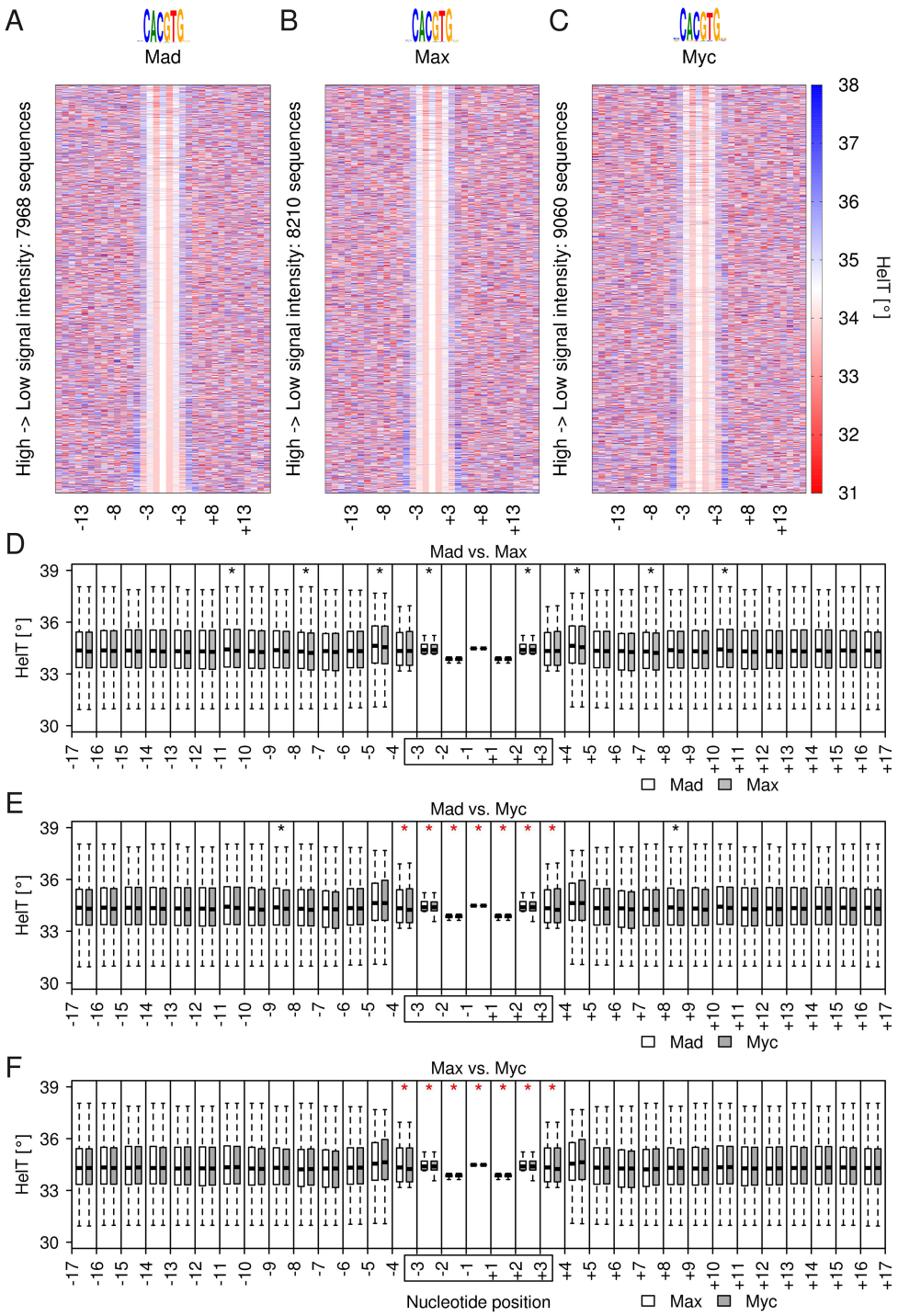
Supplementary Figure S2. DNA shape analysis of human bHLH TFBSs. (A-C) MGW preferences of the three TFs were compared using box plots. Boxes represent the median (line inside the box), 1st and 3rd quartiles (edges of the box), and the whiskers define the furthest data points within 1.5 x inter-quartile range from the edges of the box. Asterisks indicate nucleotide positions where differences in MGW distributions selected by (A) Mad vs. Max, (B) Mad vs. Myc, and (C) Max vs. Myc were significant based on a K-S test (black asterisk, $P < 0.05$; red asterisk, $P < 0.001$). (D) As a negative control, two independent experiments for Myc ('Myc1' and 'Myc2') were also compared and DNA shape selections were found to be essentially identical. MGW features were symmetrized based on the palindromic E-box.



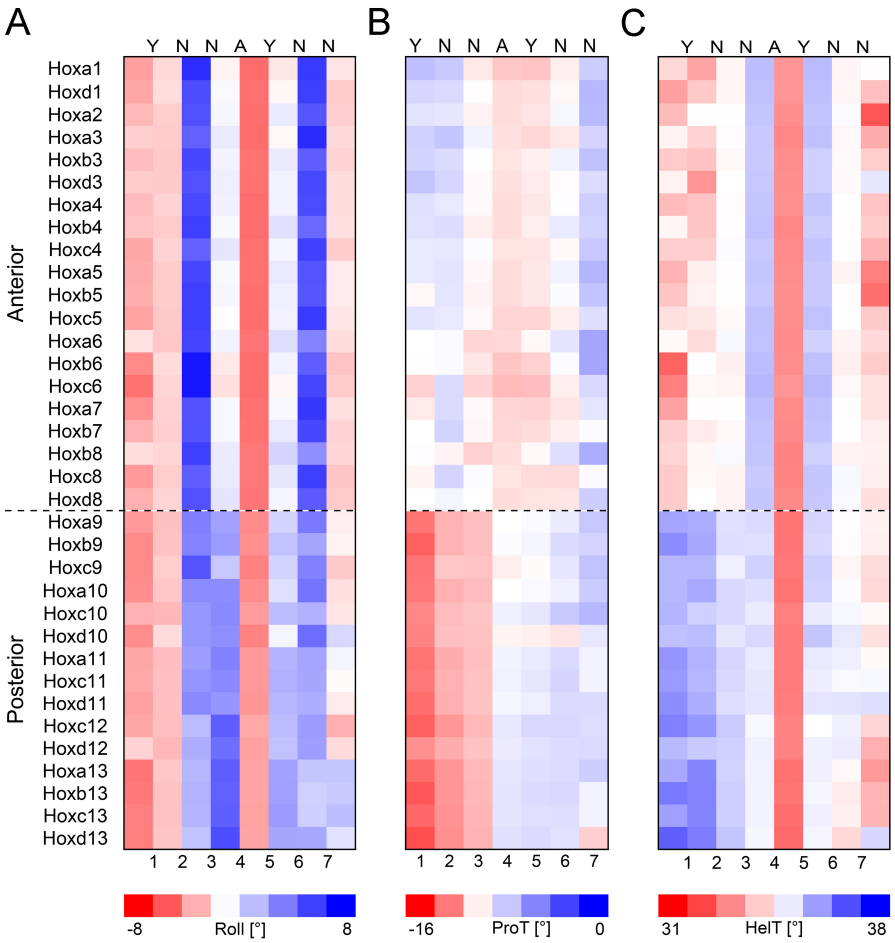
Supplementary Figure S3. DNA shape analysis of human bHLH TFBSs. Heat maps illustrate Roll selections of (A) Mad, (B) Max, and (C) Myc. (D-F) Roll preferences of the three TFs were compared using box plots. Boxes represent the median (line inside the box), 1st and 3rd quartiles (edges of the box), and the whiskers define the furthest data points within 1.5 x inter-quartile range from the edges of the box. Asterisks indicate nucleotide positions where differences in Roll distributions selected by (D) Mad vs. Max, (E) Mad vs. Myc, and (F) Max vs. Myc were significant based on a K-S test (black asterisk, $P < 0.05$; red asterisk, $P < 0.001$). Roll features were symmetrized based on the palindromic E-box.



Supplementary Figure S4. DNA shape analysis of human bHLH TFBSs. Heat maps illustrate ProT selections of (A) Mad, (B) Max, and (C) Myc. (D-F) ProT preferences of the three TFs were compared using box plots. Boxes represent the median (line inside the box), 1st and 3rd quartiles (edges of the box), and the whiskers define the furthest data points within 1.5 x inter-quartile range from the edges of the box. Asterisks indicate nucleotide positions where differences in ProT distributions selected by (D) Mad vs. Max, (E) Mad vs. Myc, and (F) Max vs. Myc were significant based on a K-S test (black asterisk, $P < 0.05$; red asterisk, $P < 0.001$). ProT features were symmetrized based on the palindromic E-box.



Supplementary Figure S5. DNA shape analysis of human bHLH TFBSs. Heat maps illustrate HelT selections of (A) Mad, (B) Max, and (C) Myc. (D-F) HelT preferences of the three TFs were compared using box plots. Boxes represent the median (line inside the box), 1st and 3rd quartiles (edges of the box), and the whiskers define the furthest data points within 1.5 x inter-quartile range from the edges of the box. Asterisks indicate nucleotide positions where differences in HelT distributions selected by (D) Mad vs. Max, (E) Mad vs. Myc, and (F) Max vs. Myc were significant based on a K-S test (black asterisk, $P < 0.05$; red asterisk, $P < 0.001$). HelT features were symmetrized based on the palindromic E-box.



Supplementary Figure S6. DNA shape analysis of mouse Hox TFBSs. Heat maps illustrate (A) Roll, (B) ProT, and (C) HelT preferrences of monomeric mouse Hox TFs determined by universal PBM experiments (26).

SUPPLEMENTARY TABLE

MGW (Figure 2)

N. P.	Mad vs. Max	Mad vs. Myc	Max vs. Myc	Myc1 vs. Myc2
-16	8.73 E-01	5.55 E-01	6.50 E-01	1.00 E+00
-15	9.11 E-01	4.65 E-01	4.56 E-01	1.00 E+00
-14	8.02 E-01	8.40 E-01	9.80 E-01	1.00 E+00
-13	6.20 E-01	8.47 E-01	1.00 E+00	1.00 E+00
-12	9.08 E-01	4.19 E-01	9.98 E-01	1.00 E+00
-11	8.64 E-01	2.08 E-01	7.14 E-01	1.00 E+00
-10	6.28 E-01	**4.74 E-02**	6.40 E-01	1.00 E+00
-9	8.75 E-01	2.24 E-01	8.93 E-01	1.00 E+00
-8	3.07 E-01	1.06 E-01	9.11 E-01	1.00 E+00
-7	1.75 E-01	8.43 E-02	8.08 E-01	1.00 E+00
-6	7.12 E-01	**1.20 E-04**	**1.13 E-04**	9.98 E-01
-5	9.62 E-02	**2.52 E-03**	**3.19 E-03**	1.00 E+00
-4	**3.26 E-02**	**7.34 E-06**	1.26 E-01	1.00 E+00
-3	4.27 E-01	**1.38 E-10**	**9.37 E-11**	1.00 E+00
-2	**2.32 E-02**	**8.46 E-06**	**6.24 E-08**	9.99 E-01
-1	5.98 E-01	**0.00 E+00**	**0.00 E+00**	2.37 E-01
+1	5.98 E-01	**0.00 E+00**	**0.00 E+00**	2.37 E-01
+2	**2.32 E-02**	**8.46 E-06**	**6.24 E-08**	9.99 E-01
+3	4.27 E-01	**1.38 E-10**	**9.37 E-11**	1.00 E+00
+4	**3.26 E-02**	**7.34 E-06**	1.26 E-01	1.00 E+00
+5	9.62 E-02	**2.52 E-03**	**3.19 E-03**	1.00 E+00
+6	7.12 E-01	**1.20 E-04**	**1.13 E-04**	9.98 E-01
+7	1.75 E-01	8.43 E-02	8.08 E-01	1.00 E+00
+8	3.07 E-01	1.06 E-01	9.11 E-01	1.00 E+00
+9	8.75 E-01	2.24 E-01	8.93 E-01	1.00 E+00
+10	6.28 E-01	**4.74 E-02**	6.40 E-01	1.00 E+00
+11	8.64 E-01	2.08 E-01	7.14 E-01	1.00 E+00
+12	9.08 E-01	4.19 E-01	9.98 E-01	1.00 E+00
+13	6.20 E-01	8.47 E-01	1.00 E+00	1.00 E+00
+14	8.02 E-01	8.40 E-01	9.80 E-01	1.00 E+00
+15	9.11 E-01	4.65 E-01	4.56 E-01	1.00 E+00
+16	8.73 E-01	5.55 E-01	6.50 E-01	1.00 E+00

Roll (Supplementary Figure 3)

N. P.	Mad vs. Max	Mad vs. Myc	Max vs. Myc	Myc1 vs. Myc2
-17|-16	8.93 E-01	4.24 E-01	9.72 E-01	9.84 E-01
-16|-15	4.86 E-01	1.48 E-01	9.56 E-01	1.00 E+00
-15|-14	4.73 E-01	2.09 E-01	5.77 E-01	1.00 E+00
-14|-13	6.65 E-01	2.58 E-01	9.82 E-01	1.00 E+00
-13|-12	9.21 E-01	4.53 E-01	9.80 E-01	9.99 E-01
-12|-11	3.52 E-01	6.56 E-02	9.76 E-01	1.00 E+00
-11|-10	2.62 E-01	**4.11 E-02**	4.56 E-01	1.00 E+00
-10|-9	2.39 E-01	**2.17 E-02**	8.32 E-01	9.92 E-01
-9|-8	1.35 E-01	**1.25 E-02**	8.62 E-01	1.00 E+00
-8|-7	1.32 E-01	6.14 E-02	1.00 E+00	1.00 E+00
-7|-6	5.04 E-02	**2.82 E-05**	1.31 E-01	1.00 E+00
-6|-5	1.14 E-01	**3.99 E-05**	1.05 E-01	1.00 E+00
-5|-4	**6.24 E-03**	**3.60 E-02**	**1.74 E-05**	9.55 E-01
-4|-3	**1.59 E-02**	6.28 E-02	8.56 E-02	1.00 E+00
-3|-2	**6.62 E-03**	**0.00 E+00**	**4.44 E-16**	9.15 E-01
-2|-1	7.69 E-01	**0.00 E+00**	**0.00 E+00**	6.89 E-01
-1|+1	5.57 E-01	**0.00 E+00**	**9.55 E-15**	2.30 E-01
+1|+2	7.69 E-01	**0.00 E+00**	**0.00 E+00**	6.89 E-01
+2|+3	**6.62 E-03**	**0.00 E+00**	**4.44 E-16**	9.15 E-01
+3|+4	**1.59 E-02**	6.28 E-02	8.56 E-02	1.00 E+00
+4|+5	**6.24 E-03**	**3.60 E-02**	**1.74 E-05**	9.55 E-01
+5|+6	1.14 E-01	**3.99 E-05**	1.05 E-01	1.00 E+00
+6|+7	5.04 E-02	**2.82 E-05**	1.31 E-01	1.00 E+00
+7|+8	1.32 E-01	6.14 E-02	1.00 E+00	1.00 E+00
+8|+9	1.35 E-01	**1.25 E-02**	8.62 E-01	1.00 E+00
+9|+10	2.39 E-01	**2.17 E-02**	8.32 E-01	9.92 E-01
+10|+11	2.62 E-01	**4.11 E-02**	4.56 E-01	1.00 E+00
+11|+12	3.52 E-01	6.56 E-02	9.76 E-01	1.00 E+00
+12|+13	9.21 E-01	4.53 E-01	9.80 E-01	9.99 E-01
+13|+14	6.65 E-01	2.58 E-01	9.82 E-01	1.00 E+00
+14|+15	4.73 E-01	2.09 E-01	5.77 E-01	1.00 E+00
+15|+16	4.86 E-01	1.48 E-01	9.56 E-01	1.00 E+00
+16|+17	8.93 E-01	4.24 E-01	9.72 E-01	9.84 E-01

ProT (Supplementary Figure 4)

N. P.	Mad vs. Max	Mad vs. Myc	Max vs. Myc	Myc1 vs. Myc2
-16	**2.67 E-02**	**1.31 E-02**	9.48 E-01	9.96 E-01
-15	**6.15 E-03**	**9.38 E-03**	9.99 E-01	1.00 E+00
-14	**1.62 E-02**	**1.50 E-03**	9.14 E-01	9.81 E-01
-13	**1.70 E-02**	**1.39 E-02**	8.77 E-01	9.86 E-01
-12	**6.49 E-03**	**1.28 E-03**	1.00 E+00	8.99 E-01
-11	**8.50 E-04**	**7.38 E-05**	9.85 E-01	1.00 E+00
-10	**4.50 E-05**	**4.25 E-05**	9.75 E-01	9.30 E-01
-9	**2.78 E-03**	**1.69 E-03**	1.00 E+00	1.00 E+00
-8	**9.68 E-05**	**8.65 E-05**	9.58 E-01	1.00 E+00
-7	**5.66 E-05**	**4.06 E-05**	9.63 E-01	9.78 E-01
-6	**1.70 E-04**	**1.08 E-06**	4.93 E-01	9.84 E-01
-5	**2.22 E-02**	**2.04 E-02**	5.58 E-01	1.00 E+00
-4	**4.72 E-02**	**0.00 E+00**	**3.53 E-14**	9.98 E-01
-3	**1.39 E-02**	**0.00 E+00**	**0.00 E+00**	1.00 E+00
-2	**8.79 E-03**	**8.70 E-11**	**8.28 E-06**	9.50 E-01
-1	6.26 E-01	**0.00 E+00**	**5.41 E-14**	2.59 E-01
+1	6.26 E-01	**0.00 E+00**	**5.41 E-14**	2.59 E-01
+2	**8.79 E-03**	**8.70 E-11**	**8.28 E-06**	9.50 E-01
+3	**1.39 E-02**	**0.00 E+00**	**0.00 E+00**	1.00 E+00
+4	**4.72 E-02**	**0.00 E+00**	**3.53 E-14**	9.98 E-01
+5	**2.22 E-02**	**2.04 E-02**	5.58 E-01	1.00 E+00
+6	**1.70 E-04**	**1.08 E-06**	4.93 E-01	9.84 E-01
+7	**5.66 E-05**	**4.06 E-05**	9.63 E-01	9.78 E-01
+8	**9.68 E-05**	**8.65 E-05**	9.58 E-01	1.00 E+00
+9	**2.78 E-03**	**1.69 E-03**	1.00 E+00	1.00 E+00
+10	**4.50 E-05**	**4.25 E-05**	9.75 E-01	9.30 E-01
+11	**8.50 E-04**	**7.38 E-05**	9.85 E-01	1.00 E+00
+12	**6.49 E-03**	**1.28 E-03**	1.00 E+00	8.99 E-01
+13	**1.70 E-02**	**1.39 E-02**	8.77 E-01	9.86 E-01
+14	**1.62 E-02**	**1.50 E-03**	9.14 E-01	9.81 E-01
+15	**6.15 E-03**	**9.38 E-03**	9.99 E-01	1.00 E+00
+16	**2.67 E-02**	**1.31 E-02**	9.48 E-01	9.96 E-01

HelT (Supplementary Figure 5)

N. P.	Mad vs. Max	Mad vs. Myc	Max vs. Myc	Myc1 vs. Myc2
-17|-16	4.17 E-01	1.90 E-01	9.99 E-01	1.00 E+00
-16|-15	5.57 E-01	7.57 E-02	8.13 E-01	1.00 E+00
-15|-14	1.78 E-01	5.63 E-01	8.77 E-01	1.00 E+00
-14|-13	7.98 E-01	1.67 E-01	5.90 E-01	1.00 E+00
-13|-12	1.65 E-01	4.27 E-01	9.73 E-01	9.70 E-01
-12|-11	6.79 E-01	2.19 E-01	9.94 E-01	1.00 E+00
-11|-10	**4.92 E-02**	1.20 E-01	9.74 E-01	9.89 E-01
-10|-9	2.45 E-01	8.25 E-02	9.06 E-01	1.00 E+00
-9|-8	1.81 E-01	**1.79 E-02**	9.63 E-01	1.00 E+00
-8|-7	**3.89 E-02**	1.49 E-01	9.25 E-01	9.65 E-01
-7|-6	1.32 E-01	1.76 E-01	1.00 E+00	1.00 E+00
-6|-5	7.02 E-01	1.95 E-01	4.04 E-01	1.00 E+00
-5|-4	**1.25 E-02**	1.27 E-01	6.73 E-01	1.00 E+00
-4|-3	1.60 E-01	**3.24 E-05**	**9.13 E-09**	1.00 E+00
-3|-2	**3.79 E-03**	**9.82 E-12**	**1.26 E-06**	7.63 E-01
-2|-1	3.46 E-01	**2.28 E-08**	**3.53 E-10**	8.72 E-01
-1|+1	5.82 E-01	**0.00 E+00**	**2.91 E-13**	2.91 E-01
+1|+2	3.46 E-01	**2.28 E-08**	**3.53 E-10**	8.72 E-01
+2|+3	**3.79 E-03**	**9.82 E-12**	**1.26 E-06**	7.63 E-01
+3|+4	1.60 E-01	**3.24 E-05**	**9.13 E-09**	1.00 E+00
+4|+5	**1.25 E-02**	1.27 E-01	6.73 E-01	1.00 E+00
+5|+6	7.02 E-01	1.95 E-01	4.04 E-01	1.00 E+00
+6|+7	1.32 E-01	1.76 E-01	1.00 E+00	1.00 E+00
+7|+8	**3.89 E-02**	1.49 E-01	9.25 E-01	9.65 E-01
+8|+9	1.81 E-01	**1.79 E-02**	9.63 E-01	1.00 E+00
+9|+10	2.45 E-01	8.25 E-02	9.06 E-01	1.00 E+00
+10|+11	**4.92 E-02**	1.20 E-01	9.74 E-01	9.89 E-01
+11|+12	6.79 E-01	2.19 E-01	9.94 E-01	1.00 E+00
+12|+13	1.65 E-01	4.27 E-01	9.73 E-01	9.70 E-01
+13|+14	7.98 E-01	1.67 E-01	5.90 E-01	1.00 E+00
+14|+15	1.78 E-01	5.63 E-01	8.77 E-01	1.00 E+00
+15|+16	5.57 E-01	7.57 E-02	8.13 E-01	1.00 E+00
+16|+17	4.17 E-01	1.90 E-01	9.99 E-01	1.00 E+00

Supplementary Table S1. DNA shape analysis of human bHLH TFBSs. The table provides the Kolmogrorov-Smirnov (K-S) test *P*-values for the comparison of DNA shape selections by Mad, Max, and Myc. The central E-box location at nucleotide positions (N.P.) -3 to +3 is highlighted in gray. Statistically significant differences in the selection of DNA shape parameters MGW (Figure 3; Supplementary Figure S2), Roll (Supplementary Figure S3), ProT (Supplementary Figure S4), and HelT (Supplementary Figure S5) are highlighted in bold black for $P < 0.05$ (black asterisks in box plots) and in bold red for $P < 0.001$ (red asterisks in box plots). As a negative control, two independent experiments for Myc ('Myc1' and 'Myc2') were also compared and DNA shape selections were found to be essentially identical.

AUTHOR CONTRIBUTIONS

L.Y. developed the methodology for generating DNA shape data from TFBSs provided by JASPAR and from PBM probes provided by UniPROBE, designed and generated the TFBSshape database, and analyzed DNA binding specificity data for human bHLH TFs. T.Z. developed the methodology for high-throughput DNA shape prediction and implemented the multiple linear regression models. I.D. analyzed mouse Hox TFBSs. A.M. and W.W. provided unpublished JASPAR data, and enabled the integration of TFBSshape with JASPAR2014. R.G. gave advice on the analysis of human bHLH TFBSs and the detection of TFBSs on probes derived from UniPROBE. L.Y. and R.R. wrote the manuscript. R.R. conceived, designed, and supervised the project.