

1 **Supplementary Tables**

2 **Supplementary Table 1. Comparisons of EPSLiM models with different underlying methods.** The
 3 Area Under the Receiver Operating Characteristic (AUROC) Curves for EPSLiM models with different
 4 combinations of individual predictors integrated. Included predictors are designated with a 1 and excluded
 5 predictors are designated with a 0. The final EPSLiM model is the first row in the table with the highest
 6 score and is represented in Figures 2 & 3.

CM-ANCHOR	MoRFpred	ANCHOR	SLiMPred	SLiMPrints	AUROC
<i>1</i>	<i>1</i>	<i>0</i>	<i>1</i>	<i>1</i>	<i>0.8364</i>
1	1	1	1	1	0.8356
1	0	1	1	1	0.8309
1	0	0	1	1	0.8279
1	1	0	1	0	0.8194
1	1	1	1	0	0.8190
0	0	1	1	1	0.8128
1	0	1	1	0	0.8106
0	0	0	1	1	0.8104
0	1	1	1	1	0.8080
1	0	0	1	0	0.8061
0	1	0	1	1	0.8038
1	1	1	0	1	0.7935
1	1	0	0	1	0.7928
0	0	1	1	0	0.7835
0	0	0	1	0	0.7802
0	1	1	1	0	0.7795
1	0	1	0	1	0.7752
0	1	0	1	0	0.7722
1	1	1	0	0	0.7679
1	1	0	0	0	0.7669
1	0	0	0	1	0.7522
0	1	1	0	1	0.7399
1	0	1	0	0	0.7365

0	0	1	0	1	0.7363
0	1	0	0	1	0.7255
1	0	0	0	0	0.7036
0	1	1	0	0	0.6911
0	0	1	0	0	0.6764
0	1	0	0	0	0.6659
0	0	0	0	1	0.6209

7

8

9 **Supplementary Table 2. Performance evaluation for SLiM predictors.** The true positive rates for
 10 Short Linear Motif (SLiM) predictors are reported for individual predictors at the specified false positive
 11 rate.

	True Positive Rate											AUROC
EPSLiM	43%	44%	58%	65%	71%	76%	80%	83%	86%	89%	90%	0.8364
ANCHOR	17%	18%	28%	33%	42%	49%	55%	60%	65%	69%	73%	0.6764
CM-ANCHOR	9%	10%	23%	36%	42%	51%	59%	66%	71%	77%	81%	0.7036
MoRFpred	25%	26%	34%	41%	45%	50%	54%	59%	62%	66%	69%	0.6659
SLiMPred	33%	34%	45%	54%	60%	65%	71%	75%	78%	82%	85%	0.7802
SLiMPrints	29%	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.6209
False Positive Rate	4.5%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	

12

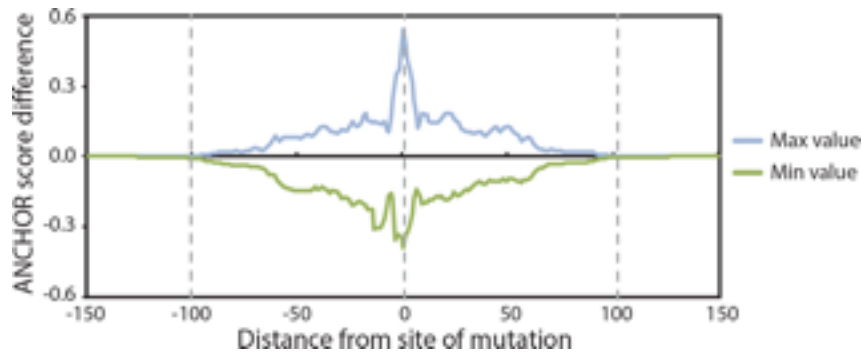
13

14 **Supplementary Table 3. Predicted binding regions and matched Eukaryotic Linear Motif (ELM)**
 15 **pattern in androgen receptor.** The sequence of human androgen receptor was input into EPSLiM and
 16 ELM database functional site predictor. 16 out of 37 EPSLiM predicted binding regions are fully or
 17 partially overlapped with at least one ELM pattern.

EPSLiM	Eukaryotic Linear Motif (ELM)
2EVQLGLGRVYPRP14	LIG_SH3_1 (9-15)
2EVQLGLGRVYPRP14	LIG_SH3_3 (9-15)
2EVQLGLGRVYPRP14	LIG_EVH1_1 (10-15)
2EVQLGLGRVYPRP14	LIG_EVH1_1 (11-15)
2EVQLGLGRVYPRP14	LIG_SH3_2 (12-17)
48APPGASLLLQQQQQ63	LIG_USP7_1 (50-54)
48APPGASLLLQQQQQ63	LIG_SUMO_SBM_2 (53-57)
74QQQQQE79	LIG_WW_Pin1_4 (78-83)
100RGPTGYLVLDEE111	LIG_FHA_1 (101-107)
100RGPTGYLVLDEE111	LIG_SH2_STAT5 (105-108)
100RGPTGYLVLDEE111	LIG_SUMO_SBM_1 (106-111)
125PERGCVPEPGA135	LIG_SH3_3 (127-133)
141KGLPQQLPAPPDEDDS156	LIG_CYCLIN_1 (141-144)
141KGLPQQLPAPPDEDDS156	LIG_SH3_3 (145-151)
141KGLPQQLPAPPDEDDS156	LIG_EVH1_1 (147-151)
141KGLPQQLPAPPDEDDS156	LIG_USP7_2 (151-157)
159P159	LIG_USP7_1 (157-161)
162LSLLGPTF169	LIG_FHA_1 (159-165)
162LSLLGPTF169	LIG_SUMO_SBM_2 (160-165)
162LSLLGPTF169	LIG_SH3_3 (164-170)
185EASTMQLLQQQQ197	LIG_FHA_1 (186-192)
2330ISDNAKEL237	LIG_FHA_2 (227-233)
317GLE319	LIG_FHA_1 (313-319)
317GLE319	LIG_SH2_STAT5 (314-317)
361DYYNFPLA368	LIG_SH2_GRB2 (362-365)
373PPPPPPPPHP381	LIG_SH3_3 (369-375)
373PPPPPPPPHP381	LIG_SH3_3 (370-376)
373PPPPPPPPHP381	LIG_SH3_3 (371-377)
373PPPPPPPPHP381	LIG_SH3_3 (372-378)
373PPPPPPPPHP381	LIG_SH3_3 (373-379)
373PPPPPPPPHP381	LIG_SH3_3 (375-381)
435TLF437	LIG_FHA_2 (436-442)
439A439	LIG_TRAF2_1 (438-441)
476AVAPYGYTRP485	LIG_SH2_STAT5 (482-485)

18

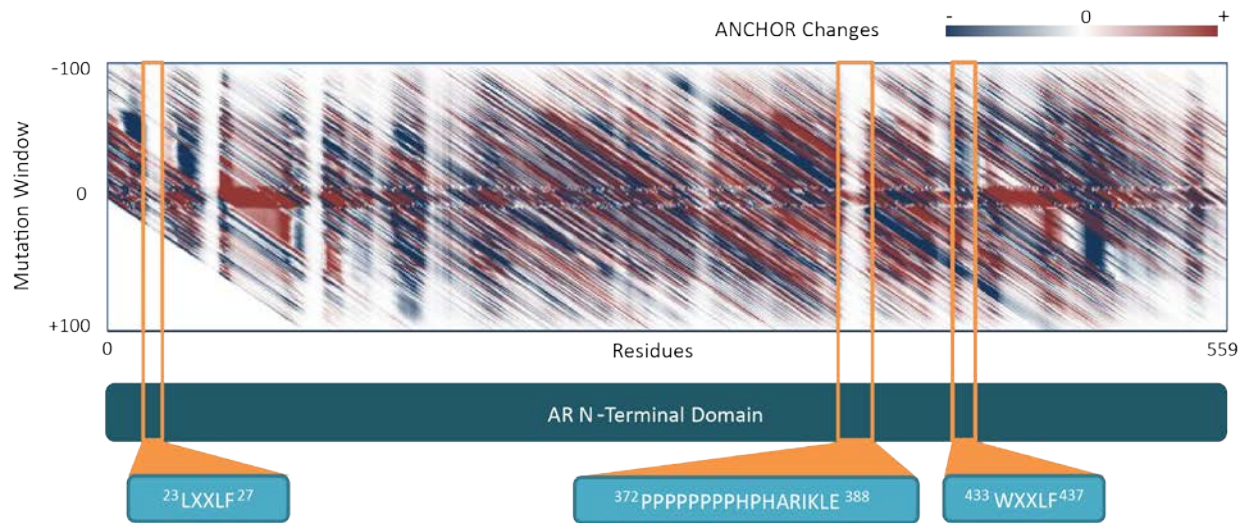
19 **Supplemental Figures.**



20

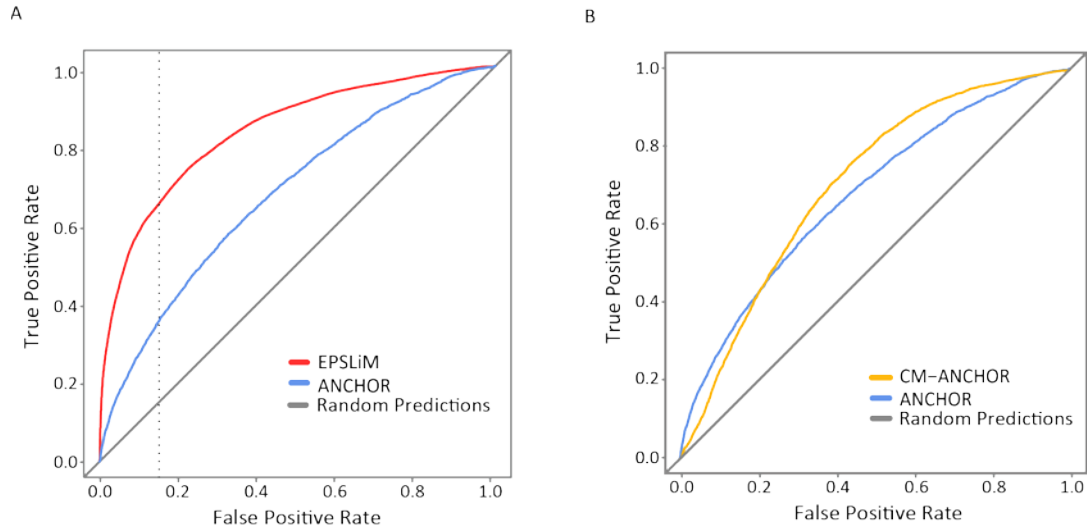
21 **Supplemental Figure 1. Distribution of ANCHOR score differences.** Mutations are introduced at each
22 residue along the disordered regions in Androgen Receptor (AR), then the binding sites are predicted
23 using ANCHOR. The score for each residue is then compared to the native AR sequence ANCHOR
24 scores. The max and min value of ANCHOR score differences are shown on the plot. The affect of a
25 mutation on ANCHOR scores does not extend past 100bp from the mutation site.

26



28 **Supplementary Figure 2. Map of ANCHOR score difference for androgen receptor N-terminal**
 29 **domain.** Each column represents the ANCHOR score differences of a residue induced by a single site
 30 substitution mutation within a window of 100 residues. The differences in ANCHOR scores are reported
 31 on a colored scale where blue represents mutations that lead to ANCHOR score decrease and red
 32 represents mutations that lead to ANCHOR score increase. Robust regions, displayed as white columns,
 33 are potential SLiM sites; several known motifs are highlighted.

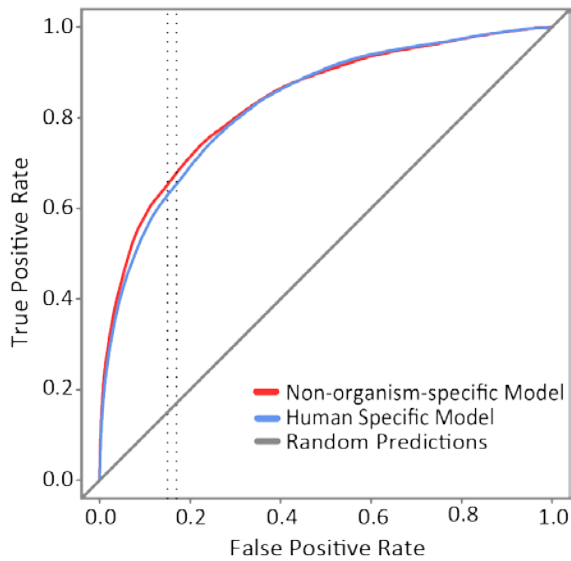
34



35

36 **Supplemental Fig. 3. Comparative analysis of ANCHOR.** ANCHOR was evaluated by measuring the
 37 Area Under the Receiver Operating Characteristic (AUROC) Curves and compared with (A) the final
 38 EPSLiM model and (B) CM-ANCHOR. For EPSLiM, AUROC = 0.8364, while ANCHOR showed a
 39 much lower performance, AUROC = 0.6764. The dotted gray line represents the 15% false positive rate.
 40 CM-ANCHOR had an overall better performance, AUROC = 0.7036.

41



42

43 **Supplementary Figure 4. Performance evaluation for EPSLiM using two different training datasets.**

44 EPSLiM was trained using a human-specific dataset and a non-organism-specific dataset, then evaluated
 45 by measuring the Area Under the Receiver Operating Characteristic (AUROC) Curves. The non-
 46 organism-specific model had the slightly better performance, AUROC = 0.8364, comparing with the
 47 human-specific model, AUROC = 0.8310. The first dotted gray line represents a 15% false positive rate,
 48 which corresponds to a 65% true positive rate at a threshold of 0.02 for the non-organism-specific model.
 49 At the same threshold, the second dotted gray line represents the false positive rate (17%) of the human-
 50 specific model.