

SUPPLEMENT Table 4: Optimized models identified from the expanded PSL

Rank	Model Name (TFs that it contains are placed in brackets) ¹	Number of gene promoters recognized in the final Problem Specific List	Number of recognized promoters in the mouse genome (69108 promoters)	Selectivity (in %)	Recall (in %)	Quality Score
1	opt2_mus_oct_ng_sox_4el_1 (V\$MAZF, V\$SORY, V\$CREB, V\$CMYB)	5	31	0.0449	35.7	0.00126
2	opt2_mus_oct_ng_sox_4el_1_3el (V\$MAZF, V\$SORY, V\$CMYB)	7	211	0.305	50.0	0.00610
3	opt2_vertical_oct_4el_04 (V\$SORY, V\$ETSF, V\$MAZF, V\$ETSF)	5	325	0.470	35.7	0.0131
4	opt3_mus_oct_ng_sox_3el_02 (V\$SORY, V\$MZF1, V\$CMYB)	5	360	0.521	35.7	0.0146
5	opt2_vertical_sox_3el_4 (V\$SORY, V\$ETSF, V\$OCT1)	6	838	1.21	42.7	0.0284
6	opt3_vertical2_Pou5f1_3el_2 (V\$MAZF, V\$ZBPF, V\$EGRF)	5	>2000	>2.8	35.7	>0.08

¹ TF binding sites are placed in the order in which they appear in the model, the first one being the most distal from the TSS (Transcription Start Site) and the last one the closest to the TSS. Framework models received a quality score that was essentially the ratio of *selectivity* versus *recall* - where *selectivity* is the percentage of all mouse promoters recognized by the model and *recall* is the percentage of promoters of input genes recognized by the model.