

Supplementary Data

TileProbe: modeling tiling array probe effects using publicly available data

Jennifer Toolan Judy¹ and Hongkai Ji^{2,*}

¹Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, 624 North Broadway, Baltimore, MD 21205, USA

²Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205, USA

Table S1. Summary of test data

TF	Platform	Species	GEO accession	Note
Gli3	Affymetrix Mouse Promoter 1.0R	Mouse	GSE11062	
Myc	Affymetrix Mouse Promoter 1.0R	Mouse	GSE11329	biotinylation mediated Myc
ER	Affymetrix Human Tiling 2.0R Array 6	Human	GSE10800	
NRSF	Affymetrix Human Tiling 2.0R Array 6	Human	GSE8489	

TF: transcription factor

Table S2. Motifs used in the analysis

TF	Source	Accession No.
Gli3	Vokes et al. (2007)	Recovered by <i>de novo</i> motif discovery
Myc	TRANSFAC (Matys et al., 2006)	M00799
ER	Carrol et al. (2005)	Recovered by <i>de novo</i> motif discovery
NRSF	TRANSFAC (Matys et al., 2006)	M00256

TF: transcription factor

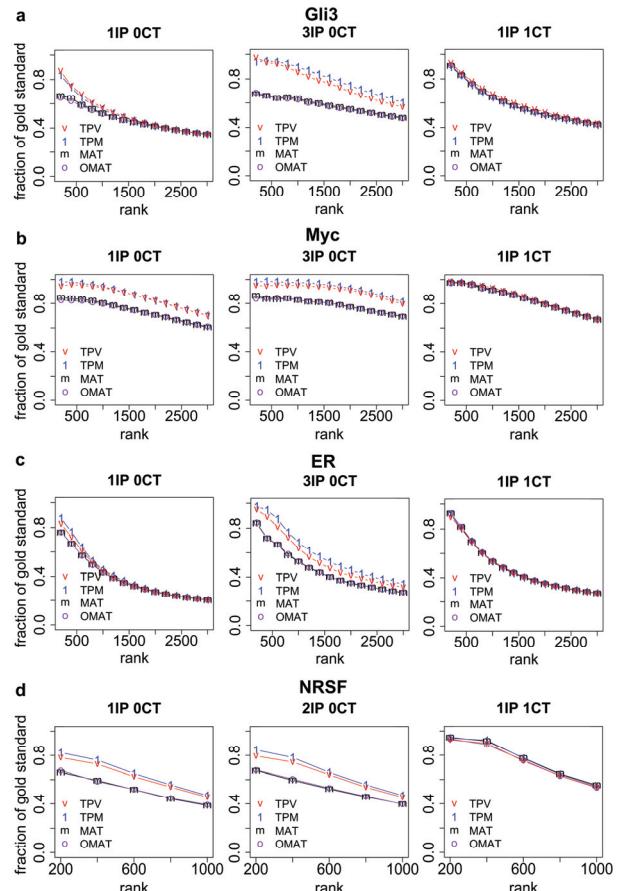


Fig. S1. Consistency test for TileProbe and MAT. The fraction of predictions that are gold standard is shown for top 200, 400, 600, ..., etc. peaks. The gold standard was constructed using the original MAT (Johnson et al., 2006) 3IP 3CT analysis. The original MAT was downloaded from <http://liulab.dfci.harvard.edu/MAT/>. OMAT: original MAT; MAT: MAT reimplemented in TileProbe. This figure shows that the original MAT and the MAT reimplemented in TileProbe are essentially the same.

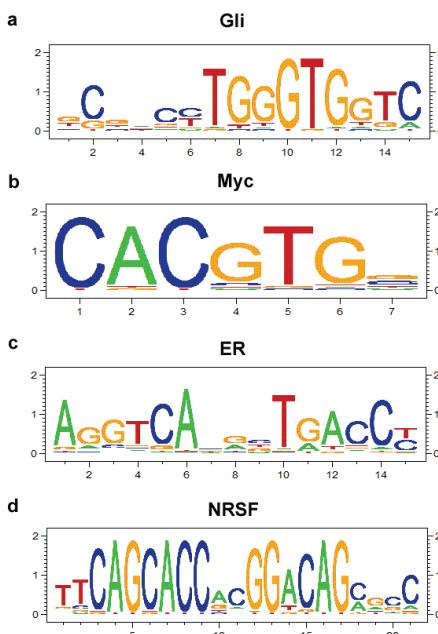


Fig. S2. Sequence logos of motifs used in the analysis.

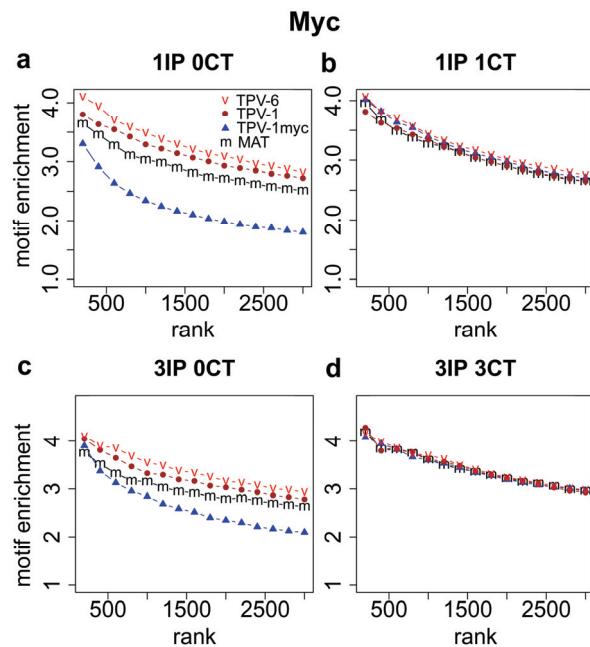


Fig. S3. Potential negative effect of using a limited number of samples to build TileProbe model. The Myc ChIP-chip data set in Table S1 was analyzed using several different probe models. TPV-6: TileProbe model was built using 6 independent studies (75 samples). TPV-1: the model was built using 1 study (6 samples) in which Myc was not involved. TPV-1myc: the model was built using 1 study (6 samples), but the study was a Myc ChIP-chip experiment (3IP 3CT) which was independent of the Myc data to be analyzed. The Myc motif enrichment among the top 200, 400, 600, ..., etc. peaks were shown. Clearly, when the training data and test data involve the

same transcription factor, TileProbe may perform worse than MAT. This illustrates that it is necessary to include a large number of samples of diverse origin to obtain a robust TileProbe model.

REFERENCES

- Carroll, J.S. et al. (2005) Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*, **122**, 33–43.
- Matys, V. et al. (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*, **34**, D108–110.
- Vokes, S.A. et al. (2007) Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning. *Development*, **134**, 1977–1989.