Supplementary - Reproducibility of parameter learning with missing observations in naive Wnt Bayesian network trained on colorectal cancer samples and Doxycycline treated cell lines[†,‡]

Shriprakash Sinha[*a]

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
First published on the web Xth XXXXXXXXXX 200X
DOI: 10.1039/b000000x

Supplementary
Contains figures and tables from the main document in Sinha[1].

[†] Electronic Supplementary Information (ESI) available: Supplementary material containing figures and tables. See DOI: 10.1039/c5mb00117j

[‡] This manuscript is a part of the Netherlands Bioinformatics Centre (NBIC) BioRange-II project BR2.5 (code - IGD71G). Matlab code is available on request for academic use only.

[a] *Netherlands Bioinformatics Centre, 6500 HB, Nijmegen, The Netherlands. Netherlands Bioinformatics Centre is supported by the Netherlands Genomics Initiative (NGI) via the Netherlands Genomics Institute in The Netherlands.*

[*] Shriprakash Sinha worked on the project while being funded by Netherlands Bioinformatics Centre. Current Address: 104-Madhurisha Heights Phase 1, Risali Sector, Bhilai - 490006, India. E-mail: shriprakash.sinha@gmail.com

Test datasets from GSE Omnibus Edgar *et al.*[4] and Barrett *et al.*[5]

Results on test datasets using Ref BN trained on 12 (6 each of Wnt off and Wnt on) samples from Colon Cancer Cell lines.

Datasets	No. Patients	Categories	TP	FP	TN	FN	P	R	F_1	FPR	ERR
GSE4183	53	N, AD, CRC, IBD	19	0	23	11	1	0.63333	0.77551	0	0.20755
GSE15960	18	N, AD, CRC	7	0	6	5	1	0.58333	0.73684	0	0.27778
GSE20916	145	N, AD, C, ADC	65	0	44	36	1	0.64356	0.78313	0	0.24828
GSE24795	30	RER CRC	26	0	0	4	1	0.86667	0.92857	-	0.13333

Results on test datasets using Ref BN trained on 64 (32 each of N and AD) samples from GSE8671 Sabates-Bellver *et al.*[3].

Datasets	No. Patients	Categories	TP	FP	TN	FN	P	R	F_1	FPR	ERR
GSE4183	53	N, AD, CRC, IBD	24	5	18	6	0.82759	0.8	0.81356	0.21739	0.20755
GSE15960	18	N, AD, CRC	12	0	6	0	1	1	1	0	0
GSE20916	145	N, AD, C, ADC	98	0	44	3	1	0.9703	0.98492	0	0.02069
GSE24795	30	RER CRC	26	0	0	4	1	0.86667	0.92857	-	0.13333

Table 1 Abbreviations : N - normal, AD - adenomas, CRC - colorectal cancer, IBD - inflammatory bowel diseases, C - carcinomas, ADC - adenocarcinomas, RER CRC - replication error colorectal cancer.

References

1 S. Sinha, *Molecular Biosystems*, 2015, **-**, 1–19.

2 M. Mokry, P. Hatzis, J. Schuijers, N. Lansu, F.-P. Ruzius, H. Clevers and E. Cuppen, *Nucleic acids research*, 2012, **40**, 148–158.

3 J. Sabates-Bellver, L. Van der Flier, M. de Palo, E. Cattaneo, C. Maake, H. Rehrauer, E. Laczko, M. Kurowski, J. Bujnicki, M. Menigatti *et al.*, *Molecular Cancer Research*, 2007, **5**, 1263.

4 R. Edgar, M. Domrachev and A. Lash, *Nucleic acids research*, 2002, **30**, 207–210.

5 T. Barrett, D. Troup, S. Wilhite, P. Ledoux, C. Evangelista, I. Kim, M. Tomashevsky, K. Marshall, K. Phillippy, P. Sherman *et al.*, *Nucleic acids research*, 2011, **39**, D1005–D1010.

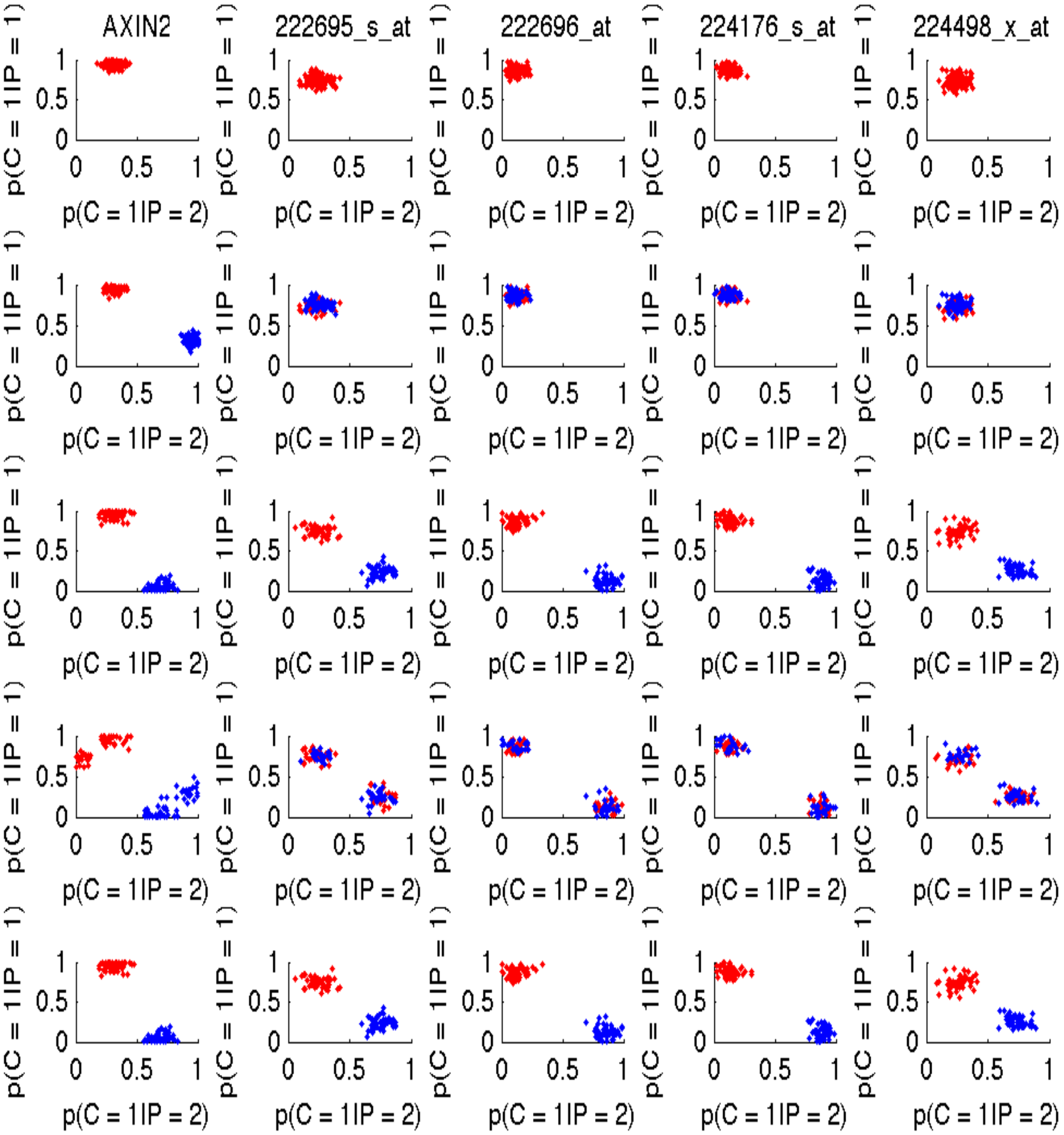


Fig. 1 Plot of deviations in estimated parameter for AXIN2 and its corresponding probesets. Rows represent the different setups. Columns indicate the node name on the top. x-axis indicates the probability of child being off given that the parent is on i.e. $p(C = 1|P = 2)$. y-axis indicates the probability of child being off given that the parent is off i.e. $p(C = 1|P = 1)$.

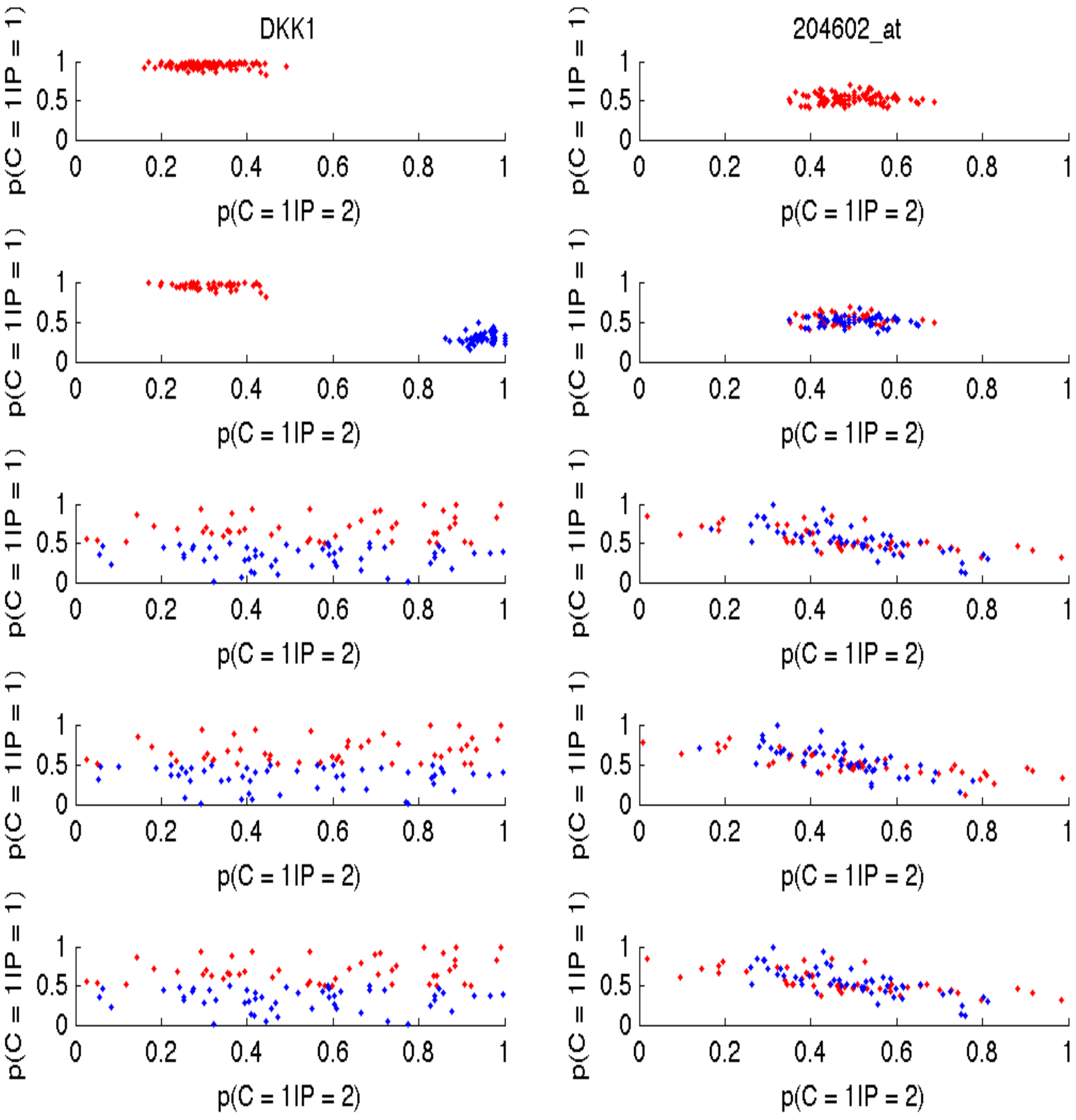
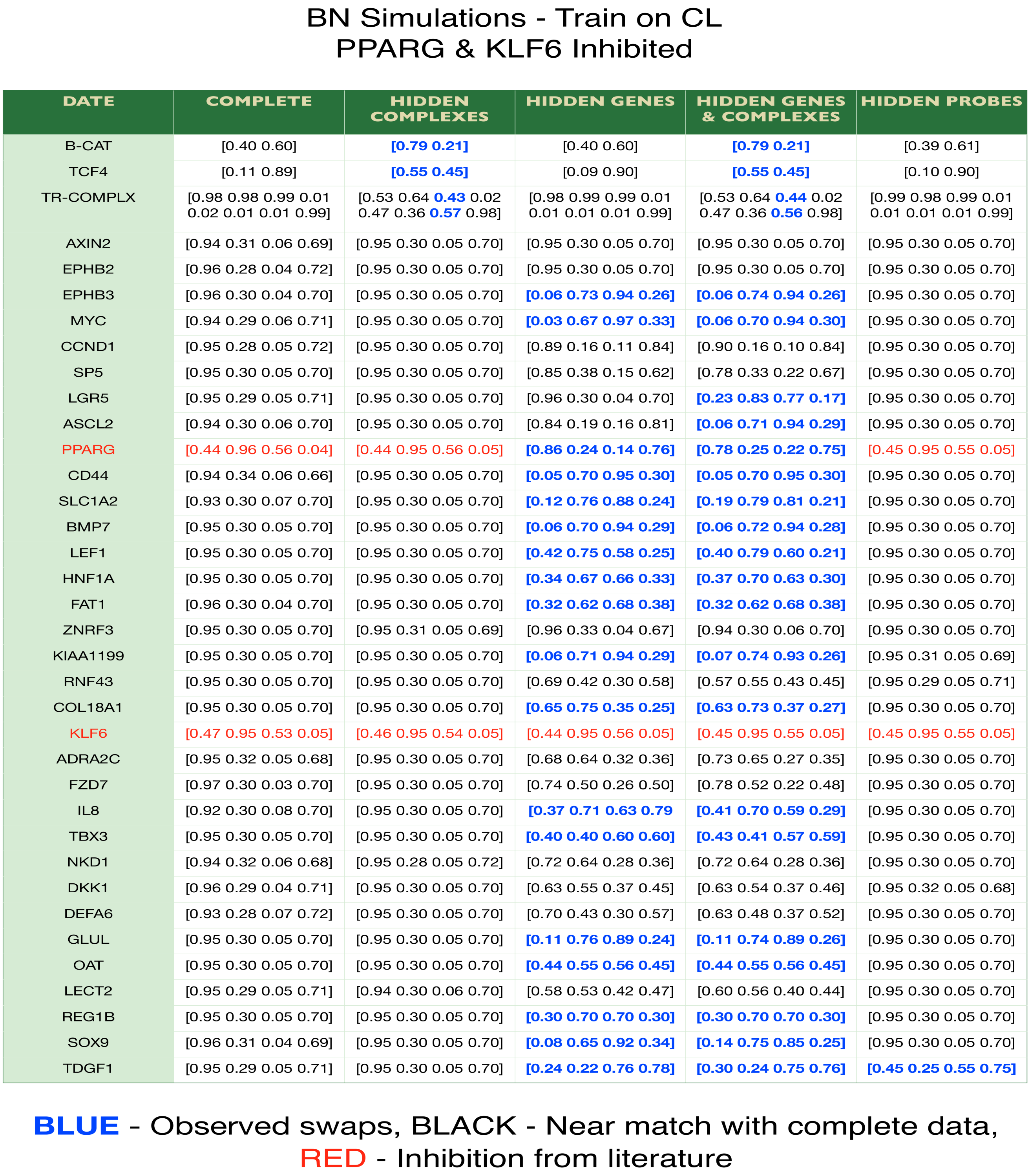


Fig. 2 Plot of deviations in estimated parameter for DKK1 and its corresponding probesets. Rows represent the different setups. Columns indicate the node name on the top. x-axis indicates the probability of child being off given that the parent is on i.e. $p(C = 1|P = 2)$. y-axis indicates the probability of child being off given that the parent is off i.e. $p(C = 1|P = 1)$.

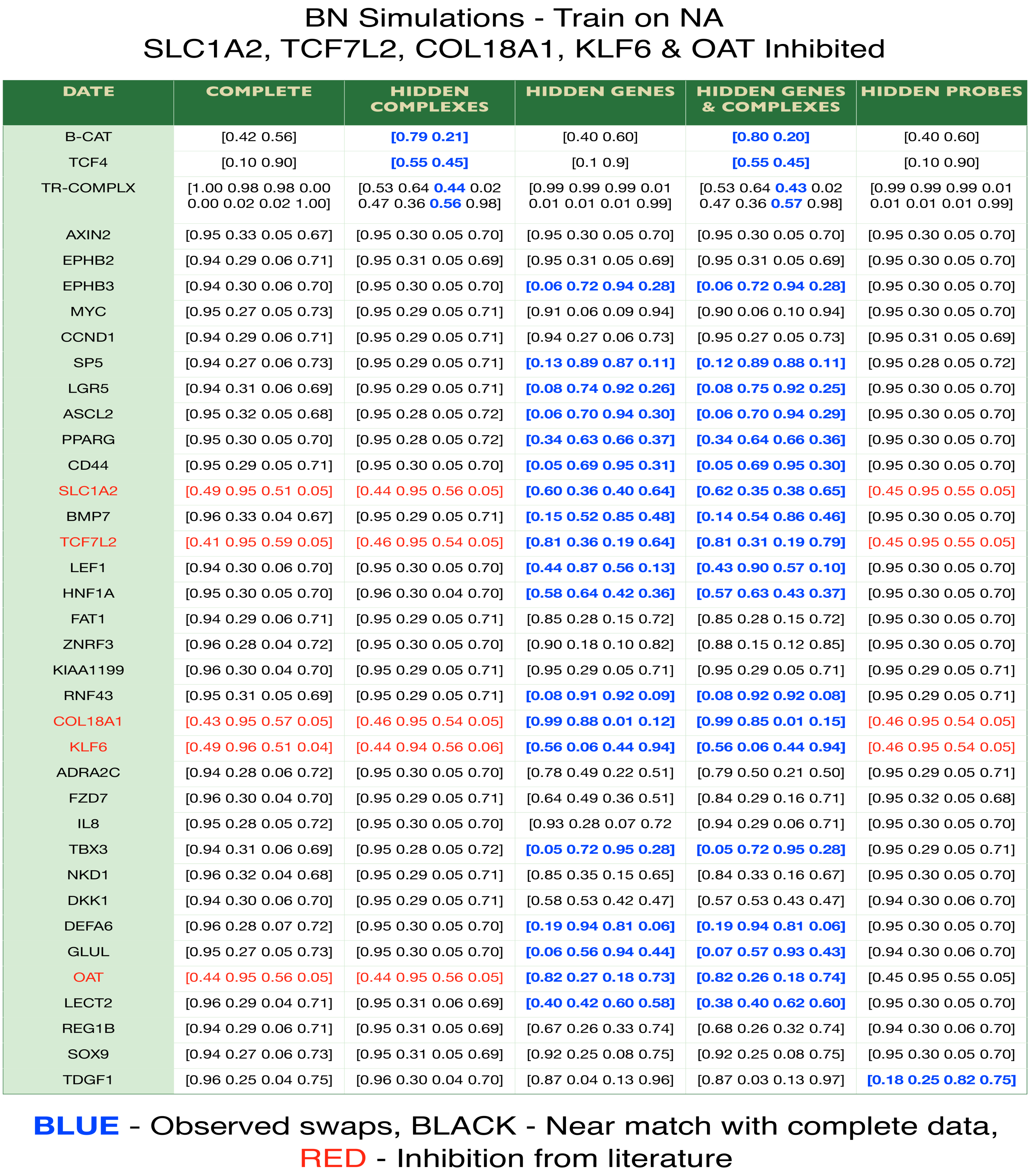


BN Simulations - Train on CL
PPARG & KLF6 Inhibited

DATE	COMPLETE	HIDDEN COMPLEXES	HIDDEN GENES	HIDDEN GENES & COMPLEXES	HIDDEN PROBES
B-CAT	[0.40 0.60]	**[0.79 0.21]**	[0.40 0.60]	**[0.79 0.21]**	[0.39 0.61]
TCF4	[0.11 0.89]	**[0.55 0.45]**	[0.09 0.90]	**[0.55 0.45]**	[0.10 0.90]
TR-COMPLX	[0.98 0.98 0.99 0.01 0.02 0.01 0.01 0.99]	[0.53 0.64 **0.43** 0.02 0.47 0.36 **0.57** 0.98]	[0.98 0.99 0.99 0.01 0.01 0.01 0.01 0.99]	[0.53 0.64 **0.44** 0.02 0.47 0.36 **0.56** 0.98]	[0.99 0.98 0.99 0.01 0.01 0.01 0.01 0.99]
AXIN2	[0.94 0.31 0.06 0.69]	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]
EPHB2	[0.96 0.28 0.04 0.72]	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]
EPHB3	[0.96 0.30 0.04 0.70]	[0.95 0.30 0.05 0.70]	**[0.06 0.73 0.94 0.26]**	**[0.06 0.74 0.94 0.26]**	[0.95 0.30 0.05 0.70]
MYC	[0.94 0.29 0.06 0.71]	[0.95 0.30 0.05 0.70]	**[0.03 0.67 0.97 0.33]**	**[0.06 0.70 0.94 0.30]**	[0.95 0.30 0.05 0.70]
CCND1	[0.95 0.28 0.05 0.72]	[0.95 0.30 0.05 0.70]	[0.89 0.16 0.11 0.84]	[0.90 0.16 0.10 0.84]	[0.95 0.30 0.05 0.70]
SP5	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	[0.85 0.38 0.15 0.62]	[0.78 0.33 0.22 0.67]	[0.95 0.30 0.05 0.70]
LGR5	[0.95 0.29 0.05 0.71]	[0.95 0.30 0.05 0.70]	[0.96 0.30 0.04 0.70]	**[0.23 0.83 0.77 0.17]**	[0.95 0.30 0.05 0.70]
ASCL2	[0.94 0.30 0.06 0.70]	[0.95 0.30 0.05 0.70]	[0.84 0.19 0.16 0.81]	**[0.06 0.71 0.94 0.29]**	[0.95 0.30 0.05 0.70]
PPARG	[0.44 0.96 0.56 0.04]	[0.44 0.95 0.56 0.05]	**[0.86 0.24 0.14 0.76]**	**[0.78 0.25 0.22 0.75]**	[0.45 0.95 0.55 0.05]
CD44	[0.94 0.34 0.06 0.66]	[0.95 0.30 0.05 0.70]	**[0.05 0.70 0.95 0.30]**	**[0.05 0.70 0.95 0.30]**	[0.95 0.30 0.05 0.70]
SLC1A2	[0.93 0.30 0.07 0.70]	[0.95 0.30 0.05 0.70]	**[0.12 0.76 0.88 0.24]**	**[0.19 0.79 0.81 0.21]**	[0.95 0.30 0.05 0.70]
BMP7	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	**[0.06 0.70 0.94 0.29]**	**[0.06 0.72 0.94 0.28]**	[0.95 0.30 0.05 0.70]
LEF1	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	**[0.42 0.75 0.58 0.25]**	**[0.40 0.79 0.60 0.21]**	[0.95 0.30 0.05 0.70]
HNF1A	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	**[0.34 0.67 0.66 0.33]**	**[0.37 0.70 0.63 0.30]**	[0.95 0.30 0.05 0.70]
FAT1	[0.96 0.30 0.04 0.70]	[0.95 0.30 0.05 0.70]	**[0.32 0.62 0.68 0.38]**	**[0.32 0.62 0.68 0.38]**	[0.95 0.30 0.05 0.70]
ZNRF3	[0.95 0.30 0.05 0.70]	[0.95 0.31 0.05 0.69]	[0.96 0.33 0.04 0.67]	[0.94 0.30 0.06 0.70]	[0.95 0.30 0.05 0.70]
KIAA1199	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	**[0.06 0.71 0.94 0.29]**	**[0.07 0.74 0.93 0.26]**	[0.95 0.31 0.05 0.69]
RNF43	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	[0.69 0.42 0.30 0.58]	[0.57 0.55 0.43 0.45]	[0.95 0.29 0.05 0.71]
COL18A1	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	**[0.65 0.75 0.35 0.25]**	**[0.63 0.73 0.37 0.27]**	[0.95 0.30 0.05 0.70]
KLF6	[0.47 0.95 0.53 0.05]	[0.46 0.95 0.54 0.05]	[0.44 0.95 0.56 0.05]	[0.45 0.95 0.55 0.05]	[0.45 0.95 0.55 0.05]
ADRA2C	[0.95 0.32 0.05 0.68]	[0.95 0.30 0.05 0.70]	[0.68 0.64 0.32 0.36]	[0.73 0.65 0.27 0.35]	[0.95 0.30 0.05 0.70]
FZD7	[0.97 0.30 0.03 0.70]	[0.95 0.30 0.05 0.70]	[0.74 0.50 0.26 0.50]	[0.78 0.52 0.22 0.48]	[0.95 0.30 0.05 0.70]
IL8	[0.92 0.30 0.08 0.70]	[0.95 0.30 0.05 0.70]	**[0.37 0.71 0.63 0.79**	**[0.41 0.70 0.59 0.29]**	[0.95 0.30 0.05 0.70]
TBX3	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	**[0.40 0.40 0.60 0.60]**	**[0.43 0.41 0.57 0.59]**	[0.95 0.30 0.05 0.70]
NKD1	[0.94 0.32 0.06 0.68]	[0.95 0.28 0.05 0.72]	[0.72 0.64 0.28 0.36]	[0.72 0.64 0.28 0.36]	[0.95 0.30 0.05 0.70]
DKK1	[0.96 0.29 0.04 0.71]	[0.95 0.30 0.05 0.70]	[0.63 0.55 0.37 0.45]	[0.63 0.54 0.37 0.46]	[0.95 0.32 0.05 0.68]
DEFA6	[0.93 0.28 0.07 0.72]	[0.95 0.30 0.05 0.70]	[0.70 0.43 0.30 0.57]	[0.63 0.48 0.37 0.52]	[0.95 0.30 0.05 0.70]
GLUL	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	**[0.11 0.76 0.89 0.24]**	**[0.11 0.74 0.89 0.26]**	[0.95 0.30 0.05 0.70]
OAT	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	**[0.44 0.55 0.56 0.45]**	**[0.44 0.55 0.56 0.45]**	[0.95 0.30 0.05 0.70]
LECT2	[0.95 0.29 0.05 0.71]	[0.94 0.30 0.06 0.70]	[0.58 0.53 0.42 0.47]	[0.60 0.56 0.40 0.44]	[0.95 0.30 0.05 0.70]
REG1B	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	**[0.30 0.70 0.70 0.30]**	**[0.30 0.70 0.70 0.30]**	[0.95 0.30 0.05 0.70]
SOX9	[0.96 0.31 0.04 0.69]	[0.95 0.30 0.05 0.70]	**[0.08 0.65 0.92 0.34]**	**[0.14 0.75 0.85 0.25]**	[0.95 0.30 0.05 0.70]
TDGF1	[0.95 0.29 0.05 0.71]	[0.95 0.30 0.05 0.70]	**[0.24 0.22 0.76 0.78]**	**[0.30 0.24 0.75 0.76]**	**[0.45 0.25 0.55 0.75]**

BLUE - Observed swaps, BLACK - Near match with complete data,
RED - Inhibition from literature

Fig. 3 Deviation of parameters using GSE18560 dataset (Mokry *et al.* [2])



BN Simulations - Train on NA
SLC1A2, TCF7L2, COL18A1, KLF6 & OAT Inhibited

DATE	COMPLETE	HIDDEN COMPLEXES	HIDDEN GENES	HIDDEN GENES & COMPLEXES	HIDDEN PROBES
B-CAT	[0.42 0.56]	**[0.79 0.21]**	[0.40 0.60]	**[0.80 0.20]**	[0.40 0.60]
TCF4	[0.10 0.90]	**[0.55 0.45]**	[0.1 0.9]	**[0.55 0.45]**	[0.10 0.90]
TR-COMPLX	[1.00 0.98 0.98 0.00 0.00 0.02 0.02 1.00]	[0.53 0.64 **0.44** 0.02 0.47 0.36 **0.56** 0.98]	[0.99 0.99 0.99 0.01 0.01 0.01 0.01 0.99]	[0.53 0.64 **0.43** 0.02 0.47 0.36 **0.57** 0.98]	[0.99 0.99 0.99 0.01 0.01 0.01 0.01 0.99]
AXIN2	[0.95 0.33 0.05 0.67]	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]	[0.95 0.30 0.05 0.70]
EPHB2	[0.94 0.29 0.06 0.71]	[0.95 0.31 0.05 0.69]	[0.95 0.31 0.05 0.69]	[0.95 0.31 0.05 0.69]	[0.95 0.30 0.05 0.70]
EPHB3	[0.94 0.30 0.06 0.70]	[0.95 0.30 0.05 0.70]	**[0.06 0.72 0.94 0.28]**	**[0.06 0.72 0.94 0.28]**	[0.95 0.30 0.05 0.70]
MYC	[0.95 0.27 0.05 0.73]	[0.95 0.29 0.05 0.71]	[0.91 0.06 0.09 0.94]	[0.90 0.06 0.10 0.94]	[0.95 0.30 0.05 0.70]
CCND1	[0.94 0.29 0.06 0.71]	[0.95 0.29 0.05 0.71]	[0.94 0.27 0.06 0.73]	[0.95 0.27 0.05 0.73]	[0.95 0.31 0.05 0.69]
SP5	[0.94 0.27 0.06 0.73]	[0.95 0.29 0.05 0.71]	**[0.13 0.89 0.87 0.11]**	**[0.12 0.89 0.88 0.11]**	[0.95 0.28 0.05 0.72]
LGR5	[0.94 0.31 0.06 0.69]	[0.95 0.29 0.05 0.71]	**[0.08 0.74 0.92 0.26]**	**[0.08 0.75 0.92 0.25]**	[0.95 0.30 0.05 0.70]
ASCL2	[0.95 0.32 0.05 0.68]	[0.95 0.28 0.05 0.72]	**[0.06 0.70 0.94 0.30]**	**[0.06 0.70 0.94 0.29]**	[0.95 0.30 0.05 0.70]
PPARG	[0.95 0.30 0.05 0.70]	[0.95 0.28 0.05 0.72]	**[0.34 0.63 0.66 0.37]**	**[0.34 0.64 0.66 0.36]**	[0.95 0.30 0.05 0.70]
CD44	[0.95 0.29 0.05 0.71]	[0.95 0.30 0.05 0.70]	**[0.05 0.69 0.95 0.31]**	**[0.05 0.69 0.95 0.30]**	[0.95 0.30 0.05 0.70]
SLC1A2	[0.49 0.95 0.51 0.05]	[0.44 0.95 0.56 0.05]	**[0.60 0.36 0.40 0.64]**	**[0.62 0.35 0.38 0.65]**	[0.45 0.95 0.55 0.05]
BMP7	[0.96 0.33 0.04 0.67]	[0.95 0.29 0.05 0.71]	**[0.15 0.52 0.85 0.48]**	**[0.14 0.54 0.86 0.46]**	[0.95 0.30 0.05 0.70]
TCF7L2	[0.41 0.95 0.59 0.05]	[0.46 0.95 0.54 0.05]	**[0.81 0.36 0.19 0.64]**	**[0.81 0.31 0.19 0.79]**	[0.45 0.95 0.55 0.05]
LEF1	[0.94 0.30 0.06 0.70]	[0.95 0.30 0.05 0.70]	**[0.44 0.87 0.56 0.13]**	**[0.43 0.90 0.57 0.10]**	[0.95 0.30 0.05 0.70]
HNF1A	[0.95 0.30 0.05 0.70]	[0.96 0.30 0.04 0.70]	**[0.58 0.64 0.42 0.36]**	**[0.57 0.63 0.43 0.37]**	[0.95 0.30 0.05 0.70]
FAT1	[0.94 0.29 0.06 0.71]	[0.95 0.29 0.05 0.71]	[0.85 0.28 0.15 0.72]	[0.85 0.28 0.15 0.72]	[0.95 0.30 0.05 0.70]
ZNRF3	[0.96 0.28 0.04 0.72]	[0.95 0.30 0.05 0.70]	[0.90 0.18 0.10 0.82]	[0.88 0.15 0.12 0.85]	[0.95 0.30 0.05 0.70]
KIAA1199	[0.96 0.30 0.04 0.70]	[0.95 0.29 0.05 0.71]	[0.95 0.29 0.05 0.71]	[0.95 0.29 0.05 0.71]	[0.95 0.29 0.05 0.71]
RNF43	[0.95 0.31 0.05 0.69]	[0.95 0.29 0.05 0.71]	**[0.08 0.91 0.92 0.09]**	**[0.08 0.92 0.92 0.08]**	[0.95 0.29 0.05 0.71]
COL18A1	[0.43 0.95 0.57 0.05]	[0.46 0.95 0.54 0.05]	**[0.99 0.88 0.01 0.12]**	**[0.99 0.85 0.01 0.15]**	[0.46 0.95 0.54 0.05]
KLF6	[0.49 0.96 0.51 0.04]	[0.44 0.94 0.56 0.06]	**[0.56 0.06 0.44 0.94]**	**[0.56 0.06 0.44 0.94]**	[0.46 0.95 0.54 0.05]
ADRA2C	[0.94 0.28 0.06 0.72]	[0.95 0.30 0.05 0.70]	[0.78 0.49 0.22 0.51]	[0.79 0.50 0.21 0.50]	[0.95 0.29 0.05 0.71]
FZD7	[0.96 0.30 0.04 0.70]	[0.95 0.29 0.05 0.71]	[0.64 0.49 0.36 0.51]	[0.84 0.29 0.16 0.71]	[0.95 0.32 0.05 0.68]
IL8	[0.95 0.28 0.05 0.72]	[0.95 0.30 0.05 0.70]	[0.93 0.28 0.07 0.72	[0.94 0.29 0.06 0.71]	[0.95 0.30 0.05 0.70]
TBX3	[0.94 0.31 0.06 0.69]	[0.95 0.28 0.05 0.72]	**[0.05 0.72 0.95 0.28]**	**[0.05 0.72 0.95 0.28]**	[0.95 0.29 0.05 0.71]
NKD1	[0.96 0.32 0.04 0.68]	[0.95 0.29 0.05 0.71]	[0.85 0.35 0.15 0.65]	[0.84 0.33 0.16 0.67]	[0.95 0.30 0.05 0.70]
DKK1	[0.94 0.30 0.06 0.70]	[0.95 0.29 0.05 0.71]	[0.58 0.53 0.42 0.47]	[0.57 0.53 0.43 0.47]	[0.94 0.30 0.06 0.70]
DEFA6	[0.96 0.28 0.07 0.72]	[0.95 0.30 0.05 0.70]	**[0.19 0.94 0.81 0.06]**	**[0.19 0.94 0.81 0.06]**	[0.95 0.30 0.05 0.70]
GLUL	[0.95 0.27 0.05 0.73]	[0.95 0.30 0.05 0.70]	**[0.06 0.56 0.94 0.44]**	**[0.07 0.57 0.93 0.43]**	[0.94 0.30 0.06 0.70]
OAT	[0.44 0.95 0.56 0.05]	[0.44 0.95 0.56 0.05]	**[0.82 0.27 0.18 0.73]**	**[0.82 0.26 0.18 0.74]**	[0.45 0.95 0.55 0.05]
LECT2	[0.96 0.29 0.04 0.71]	[0.95 0.31 0.06 0.69]	**[0.40 0.42 0.60 0.58]**	**[0.38 0.40 0.62 0.60]**	[0.95 0.30 0.05 0.70]
REG1B	[0.94 0.29 0.06 0.71]	[0.95 0.31 0.05 0.69]	[0.67 0.26 0.33 0.74]	[0.68 0.26 0.32 0.74]	[0.94 0.30 0.06 0.70]
SOX9	[0.94 0.27 0.06 0.73]	[0.95 0.31 0.05 0.69]	[0.92 0.25 0.08 0.75]	[0.92 0.25 0.08 0.75]	[0.95 0.30 0.05 0.70]
TDGF1	[0.96 0.25 0.04 0.75]	[0.96 0.30 0.04 0.70]	[0.87 0.04 0.13 0.96]	[0.87 0.03 0.13 0.97]	**[0.18 0.25 0.82 0.75]**

BLUE - Observed swaps, BLACK - Near match with complete data,
RED - Inhibition from literature

Fig. 4 Deviation of parameters using GSE8671 (Sabates-Bellver *et al.* [3])

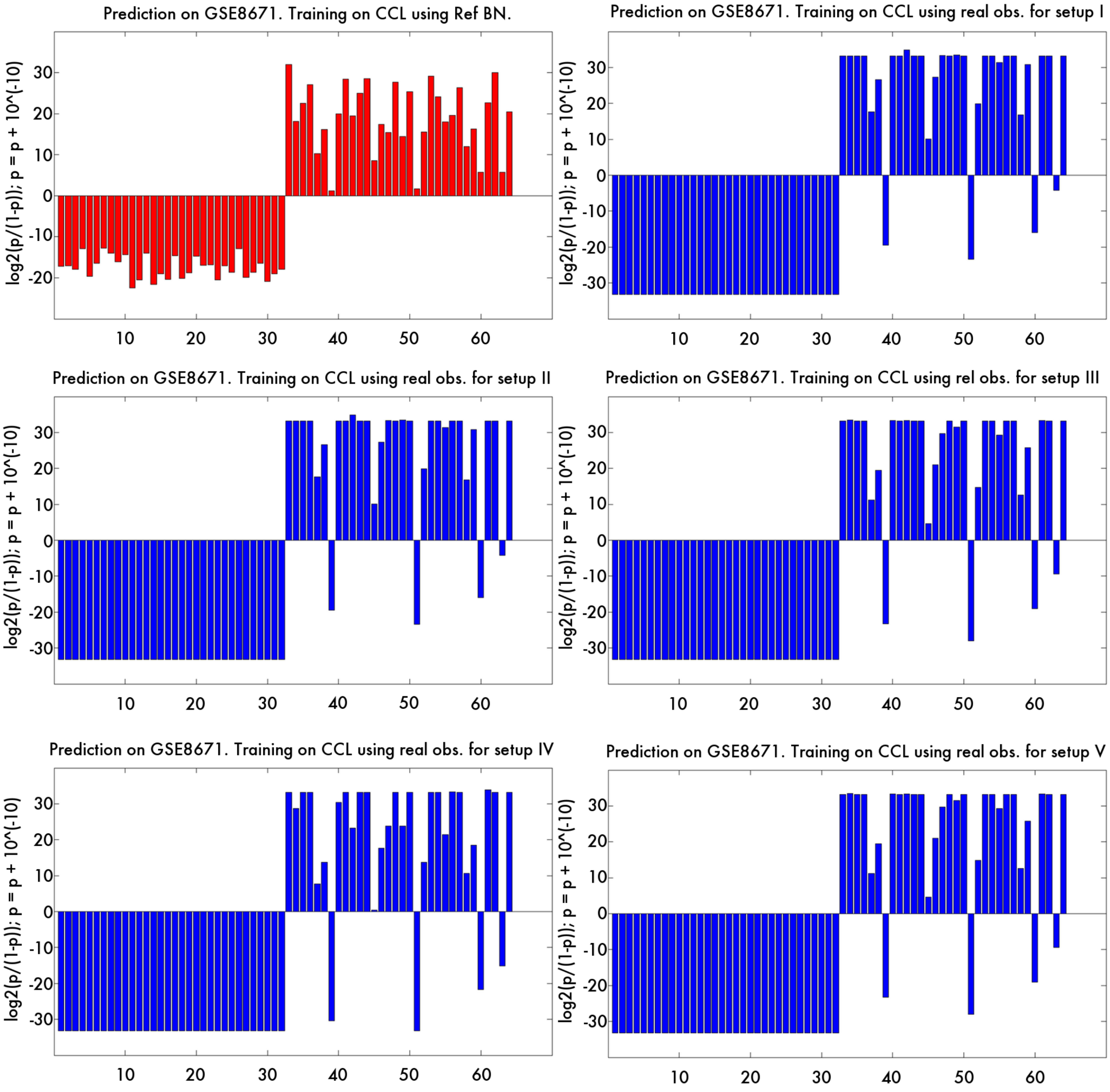


Fig. 5 Description same as in figure 5 of main manuscript. Left to right row wise, the graphs depict the predictions for BNs obtained from (a) complete data (Ref BN) (b) complete observations (BN from setup 1) (c) observed probes and genes (BN from setup 2) (d) observed probes and complexes (BN from setup 3) (e) observed probes (BN from setup 4) and (f) observed probes and $TRCMPLX$ (BN from setup 5). BNs for all setups here use real observations.

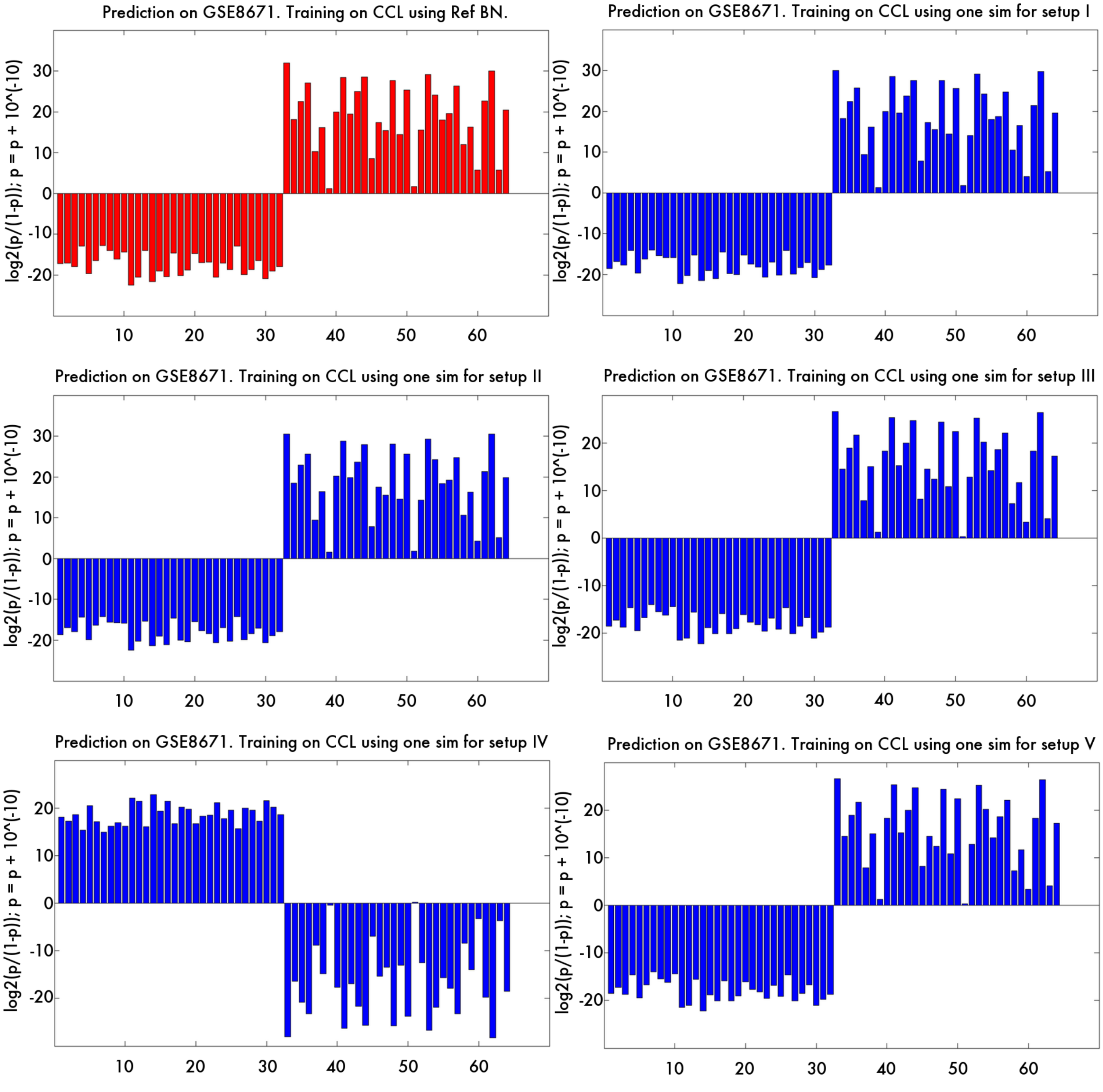


Fig. 6 Description same as in figure 5 of main manuscript. Left to right row wise, the graphs depict the predictions for BNs obtained from (a) complete data (Ref BN) (b) complete observations (BN from setup 1) (c) observed probes and genes (BN from setup 2) (d) observed probes and complexes (BN from setup 3) (e) observed probes (BN from setup 4) and (f) observed probes and $TRCMPLX$ (BN from setup 5). BNs for all setups here use sampled observations from simulation.