

Supplementary Information for Integrative Deep Models for Alternative Splicing

Anupama Jha¹, Matthew R Gazzara^{1,2,3} and Yoseph Barash^{1,2,*}

¹Department of Computer and Information Science,
School of Engineering and Applied Sciences, and

²Department of Genetics, and

³Department of Biochemistry and Biophysics,
Perelman School of Medicine,
University of Pennsylvania,
Philadelphia, PA, 19104 USA

Supplementary Table 1: Number of changing versus total events in the five tissue data used in Leung's DNN.

Events/Tissue	Brain	Heart	Kidney	Liver	Testis
Total	6570	5926	6534	5384	4716
Changing ($\Delta\text{PSII} > 0.15$)	1566	1224	1243	975	1039

Supplementary Table 2: Number of changing versus total events in the dataset 2, MGP without cassetization.

Events/Tissue	Heart	Hippocampus	Liver	Lung	Spleen	Thymus
Total	5343	5458	3898	5395	5217	6324
Changing ($\Delta\text{PSII} > 0.15$)	950	1120	614	876	721	1063

Supplementary Table 3: Number of changing versus total events in the dataset 3, MGP with cassetization.

Events/Tissue	Heart	Hippocampus	Liver	Lung	Spleen	Thymus
Total	8026	8178	5941	8234	7879	9490
Changing ($\Delta\text{PSII} > 0.15$)	1494	1752	948	1381	1145	1690

Supplementary Table 4: AUCs for inclusion vs. all, exclusion vs. all and change vs. no change for BNN-MLR and BNN-UDC

Tissue	Model	Inclusion	Exclusion	No Change
Brain-Heart	BNN-MLR	76.54 \pm 0.05	86.28 \pm 0.26	82.38 \pm 0.11
	BNN-UDC	78.83\pm0.06	87.52\pm0.09	85.45\pm0.03
Brain-Kidney	BNN-MLR	71.56 \pm 0.27	87.92 \pm 0.14	80.64 \pm 0.27
	BNN-UDC	74.62\pm0.16	89.66\pm0.07	84.57\pm0.05
Brain-Liver	BNN-MLR	70.23 \pm 0.45	88.98 \pm 0.18	79.44 \pm 0.43
	BNN-UDC	72.29\pm0.06	90.93\pm0.03	83.12\pm0.11
Brain-Testis	BNN-MLR	71.31 \pm 0.26	83.58 \pm 0.02	76.98 \pm 0.07
	BNN-UDC	74.91\pm0.23	85.11\pm0.10	81.09\pm0.10
Heart-Kidney	BNN-MLR	76.17\pm0.31	88.71 \pm 0.36	84.23 \pm 0.20
	BNN-UDC	76.17\pm0.51	90.90\pm0.05	86.59\pm0.12
Heart-Liver	BNN-MLR	71.42\pm0.50	86.80 \pm 0.21	79.61 \pm 0.30
	BNN-UDC	71.35\pm0.25	88.11\pm0.05	81.94\pm0.05
Heart-Testis	BNN-MLR	76.07 \pm 0.54	81.78 \pm 0.17	77.13 \pm 0.38
	BNN-UDC	77.06\pm0.39	83.59\pm0.12	81.42\pm0.09
Kidney-Liver	BNN-MLR	72.07 \pm 0.63	75.68 \pm 1.24	72.40 \pm 0.68
	BNN-UDC	73.58\pm1.29	76.60\pm1.41	74.68\pm0.86
Kidney-Testis	BNN-MLR	76.72 \pm 0.43	72.98\pm0.29	72.28 \pm 0.15
	BNN-UDC	78.05\pm0.19	72.83\pm0.44	77.38\pm0.27
Liver-Testis	BNN-MLR	77.11 \pm 0.27	74.49\pm0.63	73.52 \pm 0.49
	BNN-UDC	78.36\pm0.44	74.66\pm0.44	78.32\pm0.33

Supplementary Table 5: Variance explained in PSI (R^2) for BNN-MLR and BNN-UDC.

Tissue	Model	Variance Explained
Brain	BNN-MLR	21.14+-0.06
	BNN-UDC	21.30+-0.06
Heart	BNN-MLR	22.84+-0.09
	BNN-UDC	22.90+-0.07
Kidney	BNN-MLR	23.75+-0.18
	BNN-UDC	24.30+-0.09
Liver	BNN-MLR	21.54+-0.10
	BNN-UDC	21.70+-0.05
Testis	BNN-MLR	21.62+-0.03
	BNN-UDC	22.21+-0.13

Supplementary Table 6: AUCs for inclusion vs. all, exclusion vs. all and change vs. no change for BNN-UDC, DNN-LMH and DNN-PSI.

Tissue	Model	Inclusion	Exclusion	No Change
Brain-heart	BNN-UDC	78.83+0.06	87.52+-0.09	85.45+-0.03
	DNN-LMH	82.08+-0.36	84.75+-0.43	85.01+-0.24
	DNN-PSI	82.98+-0.13	86.66+-0.39	85.33+-0.23
Brain-Kidney	BNN-UDC	74.62+-0.16	89.66+-0.07	84.57+-0.05
	DNN-LMH	82.14+-0.22	88.99+-0.10	86.54+-0.08
	DNN-PSI	82.82+-0.41	90.36+-0.03	87.23+-0.06
Brain-Liver	BNN-UDC	72.29+-0.06	90.93+-0.03	83.12+-0.11
	DNN-LMH	82.33+-0.45	91.03+-0.60	86.98+-0.21
	DNN-PSI	84.31+-0.39	93.29+-0.07	87.49+-0.26
Brain-Testis	BNN-UDC	74.91+-0.23	85.11+-0.10	81.09+-0.10
	DNN-LMH	78.03+-0.27	85.23+-0.32	82.99+-0.02
	DNN-PSI	80.92+-0.71	87.44+-0.22	82.78+-0.48
Heart-Kidney	BNN-UDC	76.17+-0.51	90.90+-0.05	86.59+-0.12
	DNN-LMH	79.31+-0.35	86.40+-0.52	84.36+-0.32
	DNN-PSI	81.96+-0.35	91.72+-0.38	86.98+-0.34
Heart-Liver	BNN-UDC	71.35+-0.25	88.11+-0.05	81.94+-0.05
	DNN-LMH	76.86+-1.24	86.93+-0.67	83.13+-0.49
	DNN-PSI	80.64+-1.18	91.41+-0.29	85.58+-0.59
Heart-Testis	BNN-UDC	77.06+-0.39	83.59+-0.12	81.42+-0.09
	DNN-LMH	82.87+-0.37	84.11+-0.86	84.56+-0.21
	DNN-PSI	85.25+-0.63	90.57+-0.45	86.67+-0.63
Kidney-Liver	BNN-UDC	73.58+-1.29	76.60+-1.41	74.68+-0.86
	DNN-LMH	81.02+-2.97	69.31+-6.92	78.84+-0.22
	DNN-PSI	90.19+-0.71	87.87+-0.20	91.13+-0.26
Kidney-Testis	BNN-UDC	78.05+-0.19	72.83+-0.44	77.38+-0.27
	DNN-LMH	85.92+-1.19	83.99+-0.79	86.00+-0.34
	DNN-PSI	89.24+-0.42	89.21+-0.31	88.37+-0.25
Liver-Testis	BNN-UDC	78.36+-0.44	74.66+-0.44	78.32+-0.33
	DNN-LMH	90.81+-0.69	91.56+-0.33	91.27+-0.14
	DNN-PSI	93.69+-0.38	92.70+-0.54	91.90+-0.72

Supplementary Table 7: AUCs for inclusion vs. all, exclusion vs. all and change vs. no change for BNN-UDC without and with cassetization (BNN-CAS).

Tissue	Model	Inclusion	Exclusion	No Change
Heart-Hipp	BNN-UDC	93.00+0.12	83.80+0.15	91.03+0.04
	BNN-CAS	93.73+0.23	86.60+0.46	91.76+0.18
Heart_Liver	BNN-UDC	74.51+0.53	84.02+0.24	80.69+0.07
	BNN-CAS	75.85+0.46	87.32+0.13	82.77+0.35
Heart-Lung	BNN-UDC	79.58+0.15	84.14+0.37	83.18+0.25
	BNN-CAS	83.22+0.68	86.52+0.16	85.81+0.25
Heart-Spleen	BNN-UDC	77.16+0.71	88.10+0.14	85.22+0.09
	BNN-CAS	79.74+0.34	89.44+0.47	86.56+0.38
Heart-Thymus	BNN-UDC	75.92+1.02	83.13+0.28	82.54+0.10
	BNN-CAS	79.66+0.42	84.11+0.18	83.96+0.17
Hipp-Liver	BNN-UDC	82.07+0.35	91.34+0.23	89.09+0.25
	BNN-CAS	83.62+0.21	92.85+0.12	89.84+0.09
Hipp-Lung	BNN-UDC	81.14+0.23	91.05+0.06	89.39+0.04
	BNN-CAS	83.53+0.32	91.97+0.12	90.13+0.15
Hipp-Spleen	BNN-UDC	81.79+0.11	91.74+0.12	89.62+0.04
	BNN-CAS	83.92+0.27	93.25+0.06	90.69+0.05
Hipp-Thymus	BNN-UDC	78.67+0.38	87.27+0.04	86.19+0.14
	BNN-CAS	80.86+0.28	88.56+0.29	87.00+0.20
Liver-Lung	BNN-UDC	78.23+1.13	80.24+0.79	79.65+0.89
	BNN-CAS	81.97+0.63	83.53+0.62	83.11+0.29
Liver-Spleen	BNN-UDC	83.69+0.24	86.41+0.46	86.13+0.26
	BNN-CAS	81.27+1.49	85.07+0.65	84.71+0.46
Liver-Thymus	BNN-UDC	82.60+0.48	72.43+0.89	74.78+0.48
	BNN-CAS	85.98+1.02	75.66+0.98	77.80+0.99
Lung-Spleen	BNN-UDC	74.36+0.86	79.62+0.29	76.85+0.60
	BNN-CAS	75.34+0.91	82.15+0.47	78.44+0.77
Lung-Thymus	BNN-UDC	78.69+0.30	73.65+0.61	76.07+0.57
	BNN-CAS	81.08+0.35	75.08+1.13	77.43+0.64
Spleen-Thymus	BNN-UDC	73.27+1.53	69.79+1.40	70.44+1.62
	BNN-CAS	76.58+2.54	72.42+1.65	72.79+1.36

Supplementary Table 8: AUCs for low vs. all, medium vs. all and high vs. all for BNN-CAS and BNN-CAS with CLIP data (BNN-CAS-CLIP)

Tissue	Model	Low	Medium	High
Heart	BNN-CAS	87.38+-0.07	82.45+-0.16	87.29+-0.07
	BNN-CAS-CLIP	88.00+-0.10	83.59+-0.21	87.94+-0.14
Hipp	BNN-CAS	86.56+-0.02	84.27+-0.16	86.94+-0.06
	BNN-CAS-CLIP	87.36+-0.09	84.81+-0.13	87.71+-0.07
Liver	BNN-CAS	89.10+-0.10	84.02+-0.15	89.54+-0.10
	BNN-CAS-CLIP	89.46+-0.10	83.75+-0.26	89.92+-0.14
Lung	BNN-CAS	89.48+-0.04	81.07+-0.22	89.64+-0.05
	BNN-CAS-CLIP	90.00+-0.08	81.54+-0.29	90.13+-0.04
Spleen	BNN-CAS	89.41+-0.07	83.73+-0.05	89.71+-0.05
	BNN-CAS-CLIP	89.71+-0.07	82.89+-0.15	89.94+-0.05
Thymus	BNN-CAS	89.76+-0.09	80.56+-0.11	89.99+-0.08
	BNN-CAS-CLIP	90.19+-0.05	80.74+-0.20	90.28+-0.04

Supplementary Table 9: AUCs for inclusion vs. all, exclusion vs. all and change vs. no change for BNN-CAS and BNN-CAS with CLIP data (BNN-CAS-CLIP)

Tissue	Model	Inclusion	Exclusion	No Change
Heart-Hipp	Model	Up	Down	No Change
	BNN-CAS	92.88±0.14	88.34±0.18	92.20±0.06
Heart_Liver	BNN-CAS-CLIP	92.97±0.12	88.22±0.16	92.26±0.06
	BNN-CAS	76.93±0.48	89.42±0.15	84.70±0.23
Heart-Lung	BNN-CAS-CLIP	78.09±0.49	89.38±0.24	85.13±0.15
	BNN-CAS	81.81±0.62	89.28±0.13	87.44±0.17
Heart-Spleen	BNN-CAS-CLIP	82.52±0.67	89.77±0.18	87.94±0.18
	BNN-CAS	78.93±0.56	90.53±0.14	87.09±0.20
Heart-Thymus	BNN-CAS-CLIP	79.37±0.21	91.03±0.13	87.45±0.08
	BNN-CAS	80.29±0.37	85.36±0.09	85.01±0.04
Hipp-Liver	BNN-CAS-CLIP	82.01±0.64	86.20±0.24	85.91±0.23
	BNN-CAS	82.56±0.20	93.14±0.08	89.90±0.12
Hipp-Lung	BNN-CAS-CLIP	83.33±0.08	93.16±0.02	90.32±0.07
	BNN-CAS	83.63±0.21	92.60±0.03	90.58±0.03
Hipp-Spleen	BNN-CAS-CLIP	84.19±0.23	92.71±0.05	90.61±0.04
	BNN-CAS	82.70±0.24	93.32±0.03	90.51±0.10
Hipp-Thymus	BNN-CAS-CLIP	83.84±0.34	93.36±0.06	90.75±0.10
	BNN-CAS	82.42±0.21	88.29±0.09	87.47±0.05
Liver-Lung	BNN-CAS-CLIP	83.10±0.36	88.63±0.15	87.83±0.18
	BNN-CAS	84.02±0.63	80.35±0.67	82.16±0.54
Liver-Spleen	BNN-CAS-CLIP	84.60±0.36	81.73±0.37	83.07±0.42
	BNN-CAS	85.13±0.94	87.29±0.08	87.28±0.21
Liver-Thymus	BNN-CAS-CLIP	85.41±0.40	87.66±0.15	87.59±0.21
	BNN-CAS	84.06±0.80	76.22±0.47	78.45±0.48
Lung-Spleen	BNN-CAS-CLIP	84.25±1.10	74.23±0.03	77.03±0.14
	BNN-CAS	80.08±0.90	81.15±0.11	81.06±0.52
Lung-Thymus	BNN-CAS-CLIP	79.82±0.32	80.71±0.49	80.71±0.09
	BNN-CAS	80.07±0.46	77.64±0.58	78.73±0.38
Spleen-Thymus	BNN-CAS-CLIP	79.97±0.41	78.41±0.44	79.57±0.30
	BNN-CAS	73.98±1.64	71.64±0.95	72.06±0.78

Supplementary Table 10: CLIP-Seq experiments used to derive CLIP features

RBP	Condition	Run accession numbers	Dataset accession number	PMID
Celf1	C2C12	SRR1593050- SRR1593052	GSE61893	25883322
Celf1	C2C12	ERR039244- ERR039249	E-MTAB- 414	22355723
Celf1	heart	SRR1593053- SRR1593054	GSE61893	25883322
Celf1	muscle	SRR1593055- SRR1593056	GSE61893	25883322
Celf2	N2A nueroblastoma	SRR3179956	GSE78111	27253061
Mbnl1	brain	SRR533660,SRR533661	GSE39911	22901804
Mbnl1	C2C12	SRR533665	GSE39911	22901804
Mbnl1	C2C12	ERR039250- ERR039254	E-MTAB- 414	22355723
Mbnl1	heart	SRR533662	GSE39911	22901804
Mbnl1	muscle	SRR1553158- SRR1553160	GSE60487	25263597
Mbnl1	muscle	SRR533663	GSE39911	22901804
Ptbp1	C2C12	ERR039255- ERR039256	E-MTAB- 414	22355723
Rbfox1	brain	SRR1035719	SRP030031	24213538
Rbfox2	brain	SRR1035720	SRP030031	24213538
Rbfox2	cardiomyocytes	SRR1297227- SRR1297228	GSE57926	25753418

Supplementary Table 11: RNA-Seq data for Splice Factor Knockout/Knockdown/Overexpression experiments

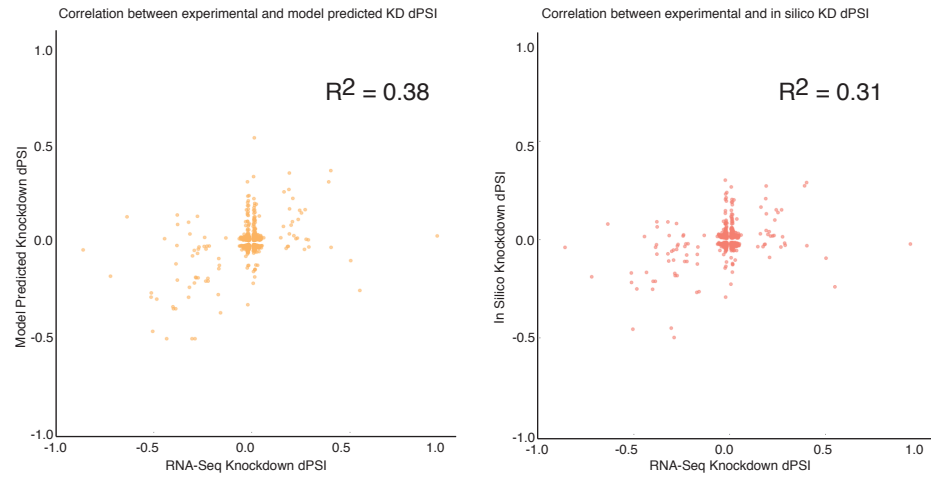
RNA-Seq Experiment	Dataset Accession	PMID
CELF1 OE heart	GSE56185	24752171
CELF2 OE heart	GSE61893	25883322
CELF1 OE muscle	GSE61893	25883322
Rbfox2 KD myotubes (differentiated C2C12)	GSE58928	25087874
Mbnl1 KO heart	GSE39911	22901804
Mbnl1 KO brain	GSE39911	22901804
Mbnl1 KO muscle	GSE39911	22901804

Supplementary Table 12: Hyperparameters were tuned for DNN-LMH and DNN-PSI using the Spearmint package similar to Leung et al, 2014

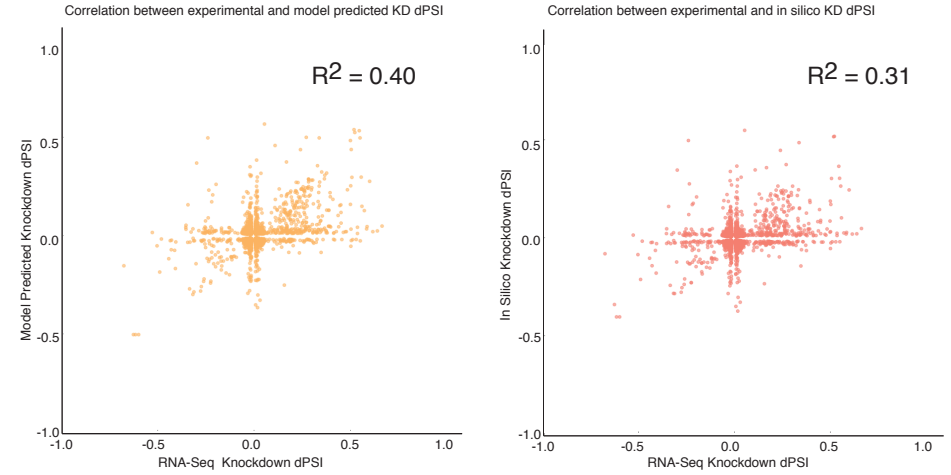
Hyperparameter	Initial Range	Final Range
Learning Rate LMH Code	0.001-0.02	0.001-0.003
Learning Rate DNI Code	0.01-0.05	0.02-0.045
Learning Rate Autoencoder	0.001-0.02	0.001-0.005
Standard deviation weights	0.01-0.05	0.01-0.03
Minibatch Size	1200-2000	1200-1400
Hidden Layer 1	300-800	500-650
Hidden Layer 2	3000-8000	6000-7000
Hidden Layer 3	450-800	500-700

Supplementary Figure 1: Title of each panel shows the splice factor KO/KD/OE in the tissue of interest. Left figure of each panel shows the correlation between RNA-Seq KD/KO/OE dPSI (x-axis) and the model predicted KD/KO/OE dPSI (y-axis). Right figure of each panel shows the correlation between RNA-Seq KD/KO/OE dPSI (x-axis) and the in silico KD dPSI (y-axis).

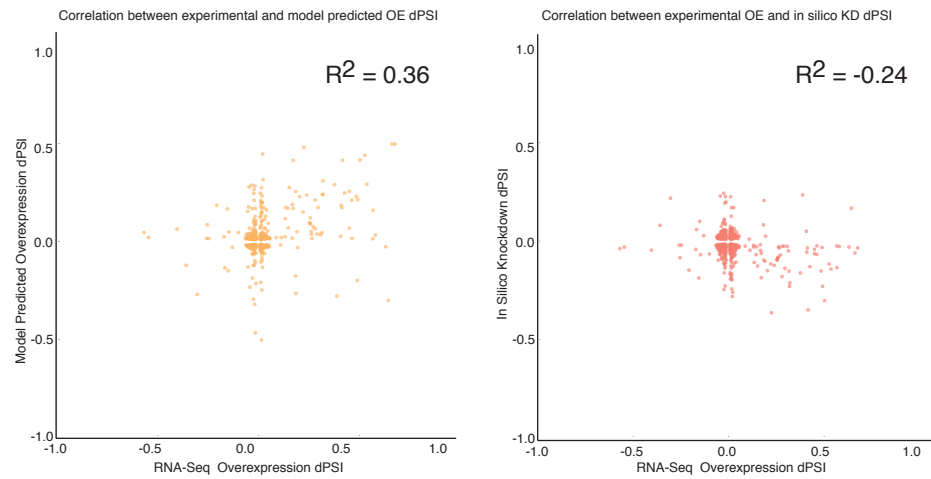
Mbn1 muscle Knockout



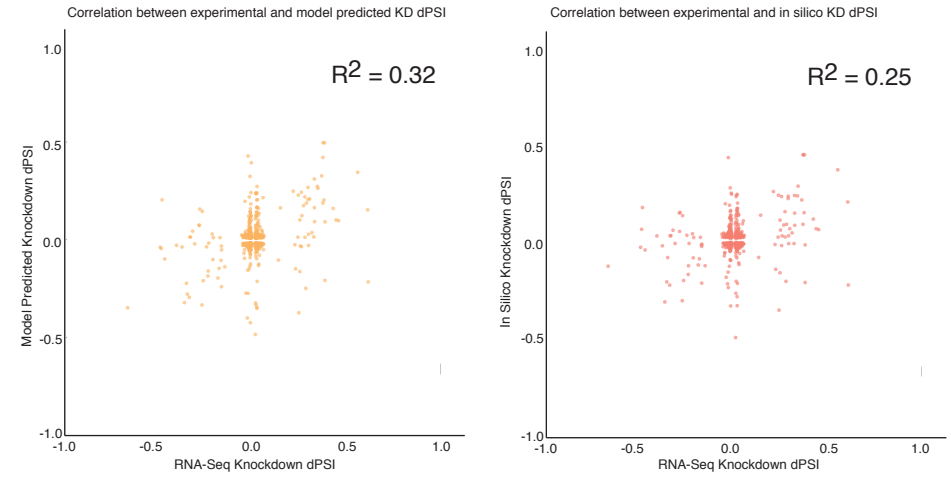
Rbfox2 Knockdown (KD) in C2C12 Cells



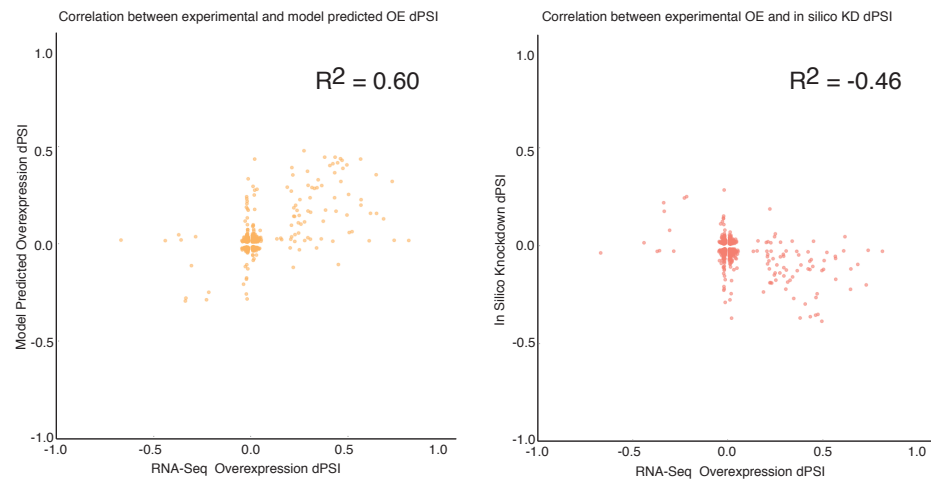
Celf2 heart Overexpression



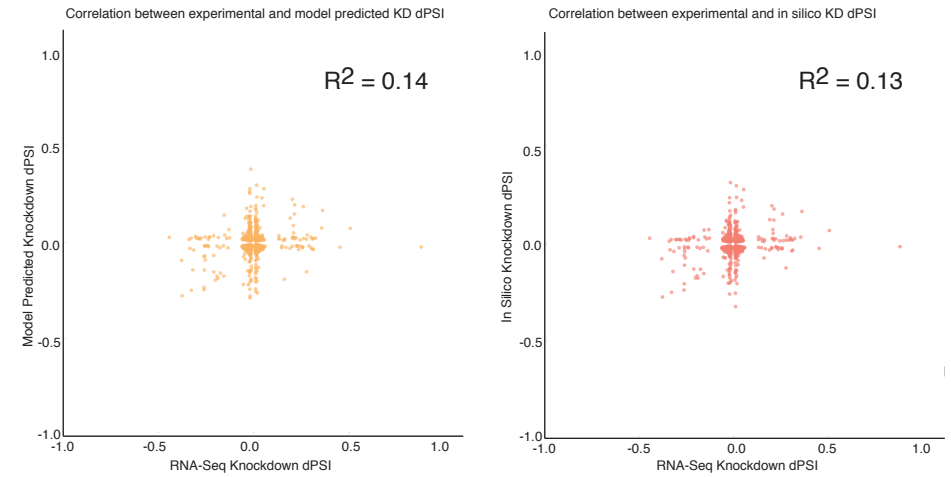
Mbn1 heart Knockout



Celf1 muscle Overexpression



Mbn1 brain Knockout



Supplementary Figure 2: Comparison of Leung et al., 2014 model and our reconstruction

