A Supplemental Information

A.1 Supplemental Data Items

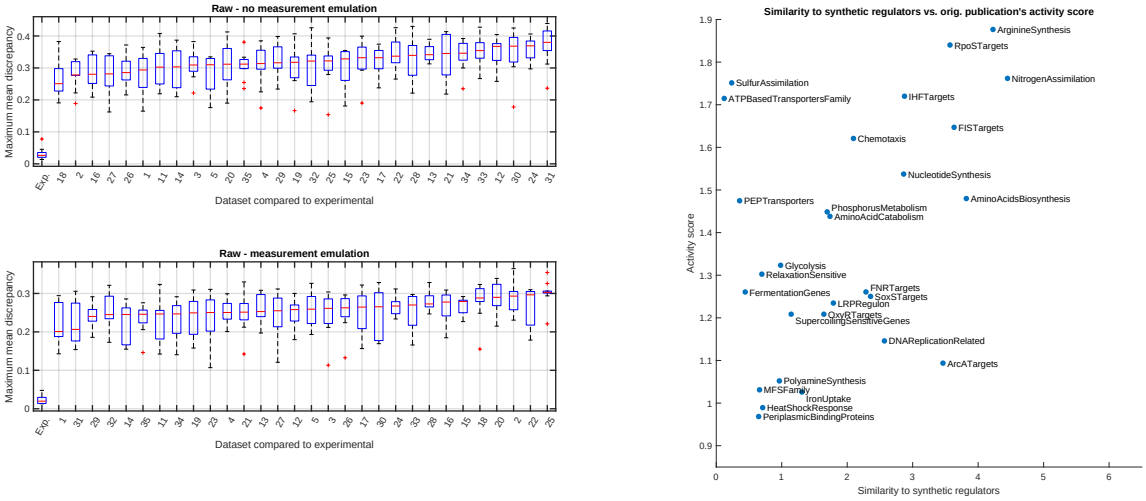

Figure S1. Related to section 2.1. (A) Maximum mean discrepancy between simulated (see Table S3) and experimental data. In both rows the raw data (non-standardized) was used for comparison, in the top row without measurement emulation; in the bottom row with additional measurement emulation. The MMD was computed according to section 4.4. (B) Similarity of experimental to synthetic data per gene class vs. the classes' activity score as reported in [53]. The similarity is the $-log_{10}$(p-value) of the enrichment of genes of the particular class among those most similar to the active, synthetic regulator/regulated pairs. (See section 4.4 for details.)

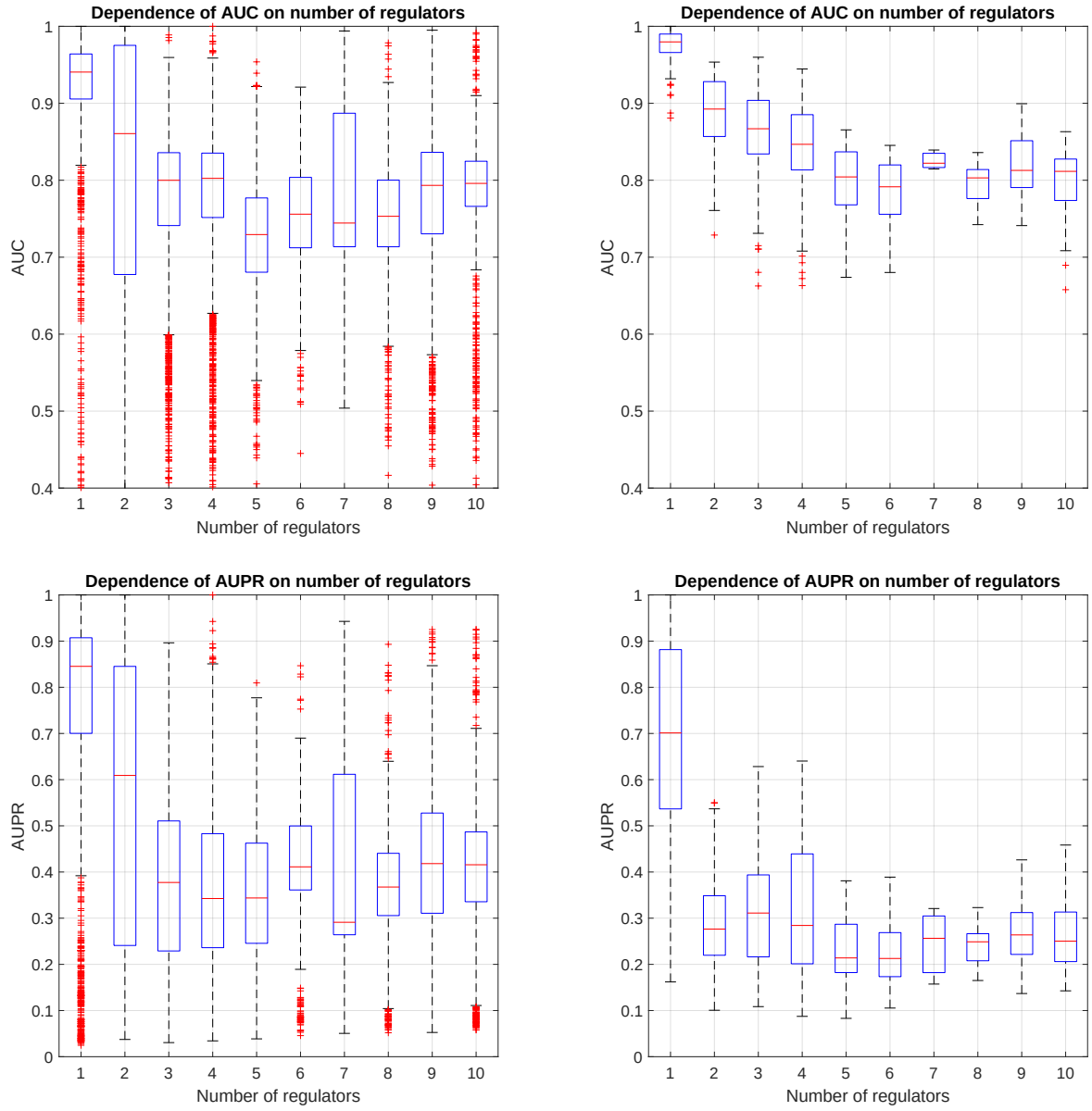

Figure S2. Related to section 2.3. Reconstruction performance dependence on type of perturbation and number of regulators for individual networks in two different test sets (columns) in terms of AUC (top row) and AUPR (bottom row). The synthetic data sets *35* (left) and *25* (right) were simulated the same way with the exception of the application of multiple gene perturbations for set *25*. Results for the *ccbld* architecture trained and predicting on each of the two sets individually. *Regulator* refers here to a gene with outdegree $\geq$ 2.

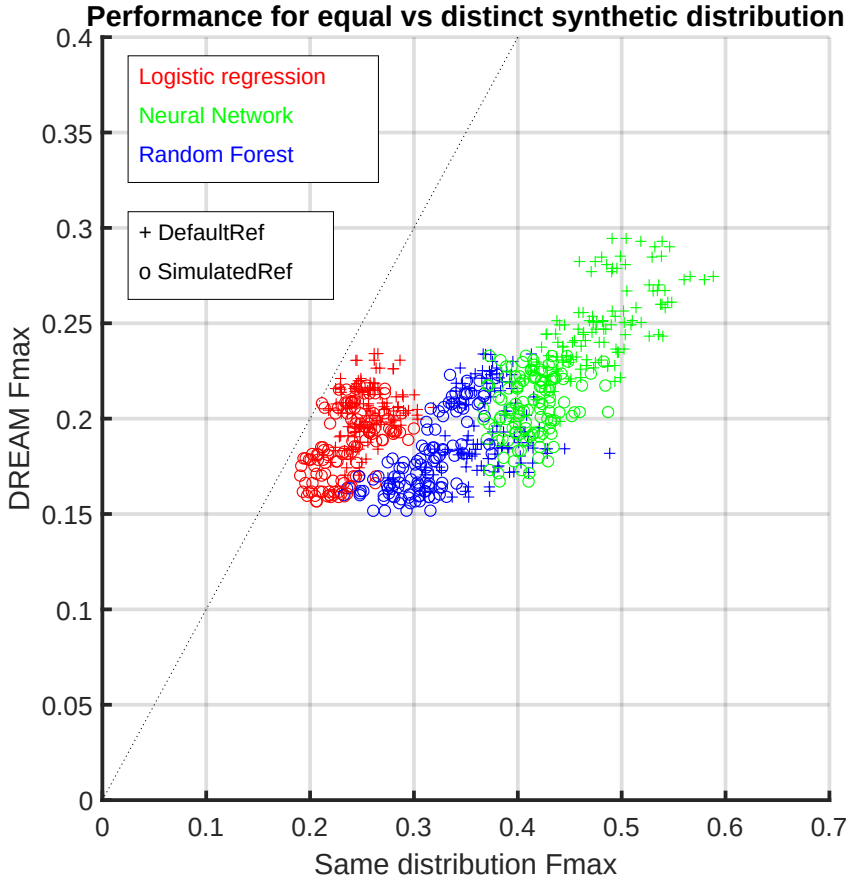

Figure S3. Related to section 2.4. Loss of reconstruction performance for classifiers predicting on the DREAM4-like data when they had been trained on distinct synthetic data sets. Original F_{max} score on the same synthetic dataset vs. F_{max} score on DREAM4-like data. Results for neural network architecture *ccbld*, random forests with depths up to 100, 150 % of default feature count and feature vector *cas*, and logistic regression for feature vector *cas.oaa*. The dashed line indicates the main diagonal.

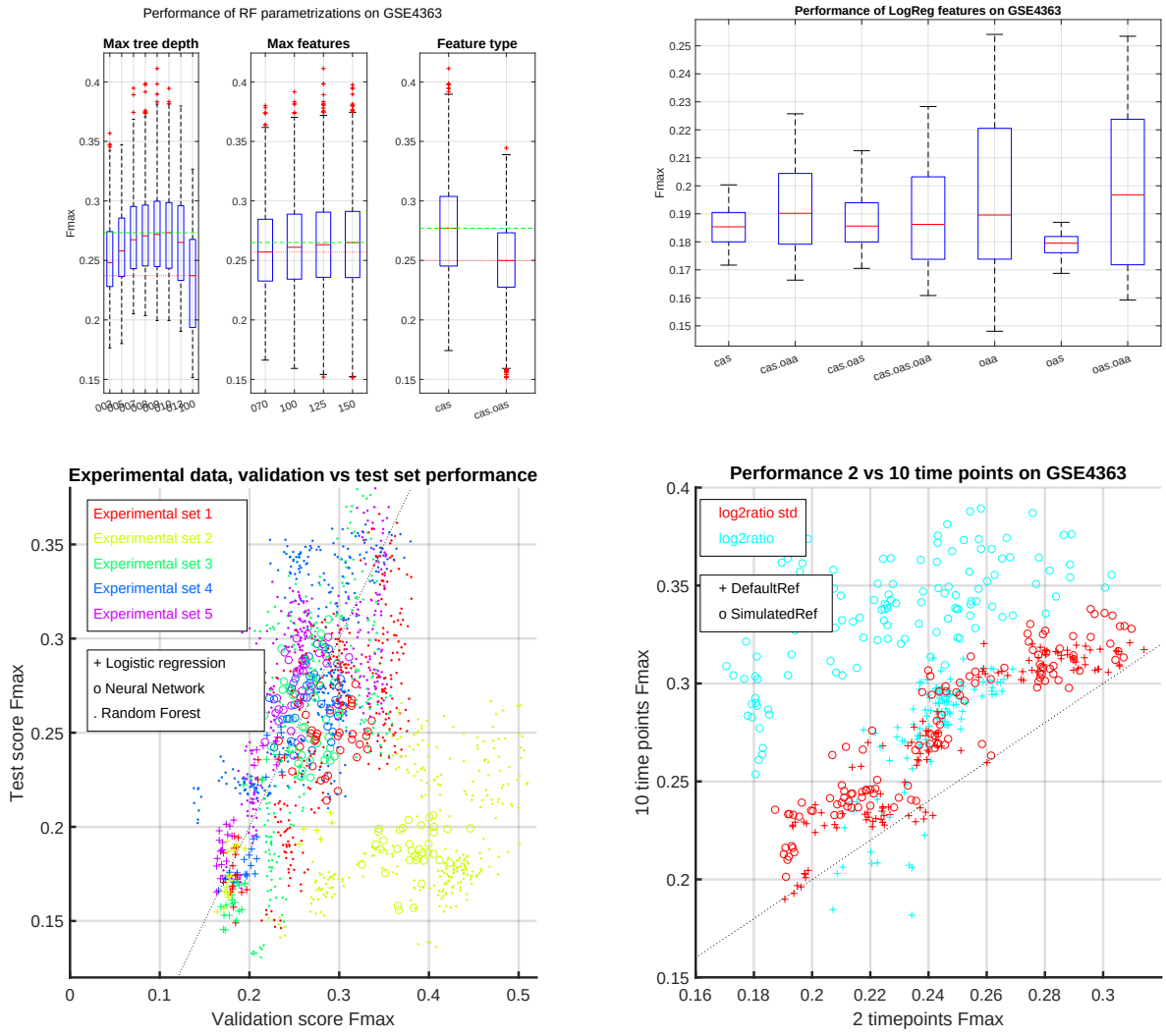

Figure S4. Related to section 2.5. (A) Reconstruction performance (F_{max}) dependence on random forest parametrization on experimental data. (Left panel) Max. tree depth. (Center panel) Max features in % of default value sqrt(| feature vector|). (Right panel) Feature vectors, *cas* concatenated time series, *cas.oas* concatenated time series plus outer product of time series. (B) Reconstruction performance (F_{max}) dependence on feature vector (see table 2) for logistic regression and experimental data. (C) Comparison reconstruction performance (F_{max}) between validation and test set for each of the five individual splits of the same experimental data set. F_{max} score for the validation set vs the corresponding test set. Results for the *selected classifiers* as described in section 2.5. The dashed line is the main diagonal. (D) Comparison of reconstruction performance (F_{max}) using only the first and last time point vs. using all ten available time points of the experimental data set. Results on the two time points for random forests (of the *selected classifiers* according to section 2.5), which were trained on only the first and last time point of each synthetic data set. The dashed line is the main diagonal.

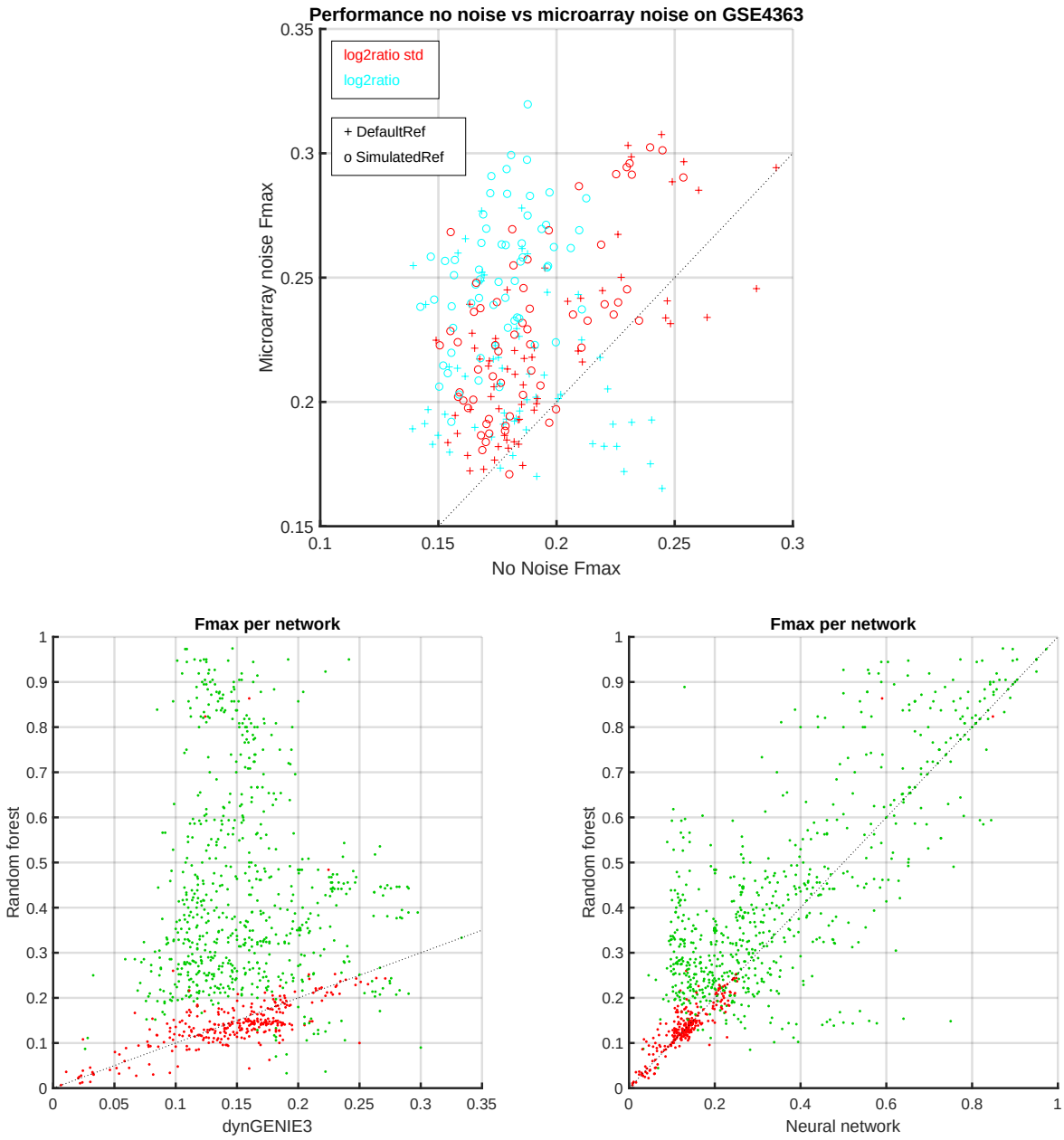

Figure S5. Relating to section 2.5. (A) Comparison of reconstruction performance (F_{max}) on the experimental data from neural networks trained on synthetic data with no noise vs. trained on the same synthetic data but with microarray noise applied (see section 4.1.5). The dashed line is the main diagonal. (B) Comparison of reconstruction performance (F_{max}) for each individual network in the experimental data split 1 for two methods. Green color indicates that the results is significant in at least one of the two compared methods. Red means both methods did not yield a significant prediction ($\alpha = 0.05$, see section 4.6 for details). (Left panel) dynGENIE3 vs. random forest. (Right panel) Neural network vs. random forest. The dashed line is the main diagonal.

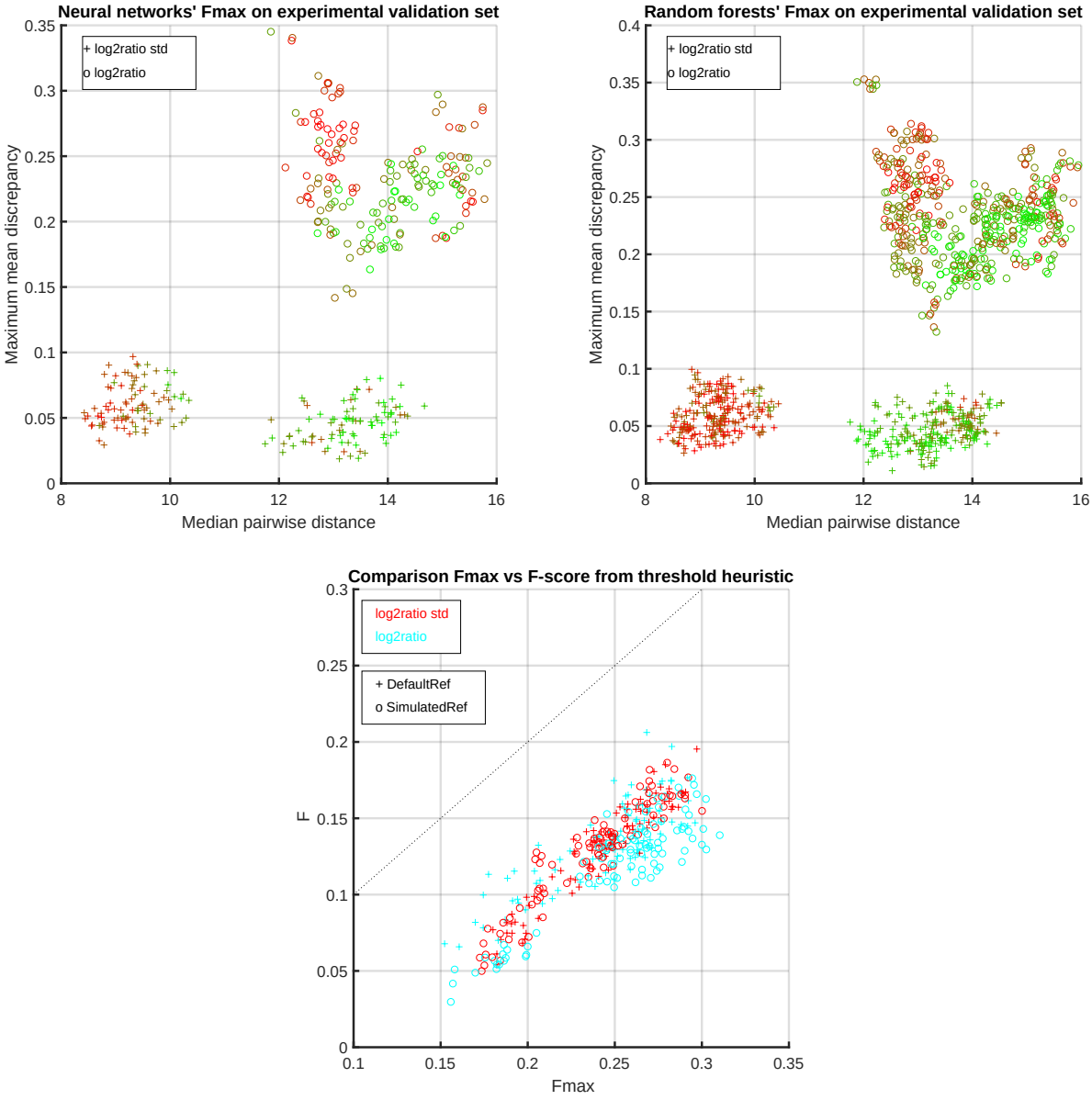

Figure S6. Related to section 3. (A) Reconstruction performance on the experimental validation sets by median pairwise distance of training set and Maximum Mean Discrepancy between training and test data. The colorcode corresponds to low (red) and high (green) values of F_{max}. (Left panel) Results for the *selected* neural network architectures (see section 2.5). (Right panel) Results for the *selected* random forest parametrizations (see section 2.5). (B) Comparison of maximally achievable (F_{max}) and heuristically determined (F) reconstruction performance on the experimental test. For F_{max} the optimal threshold is assumed to be known. For the heuristic, we compute all optimal prediction score thresholds in the validation set and apply their mean as threshold for the corresponding test set prediction scores. Results for the selected *cr* architectures. The dashed lines indicates the main diagonal.

Table S1. Setting for GeneNetWeaver's sub-sampling of networks from the Regulon-DB graph.

Setting	Value
input-net	GML file exported from Regulon-DB
min-num-regulators	20
num-subnets	1000/100/200 for synthetic train/val/test, 500 for experimental val/test
rat-selection	20
subnet-size	20

Table S2. Simulation settings for GeneNetWeaver which were changed from their default values.

Setting	Value
appendZeroInteractionsInGoldStandardFiles	0
ssKnockouts	0
ssKnockdowns	0
ssMultifactorial	0
ssDREAM4TimeSeries	0
ssDualKnockouts	0
maxtSteadyStateODE	800
maxtSteadyStateSDE	800
tsMultifactorial	1
tsDREAM4TimeSeries	0
maxtTimeSeries	400
dt	5
loadPerturbations	1
timeStepSDE	0.1
noiseCoefficientSDE	0
addMicroarrayNoise	1
normalizeAfterAddingNoise	0

Table S3. Configurations for synthetic dataset generation. *Perturbation signal type* shows the probabilities for randomly choosing any of the four types. *Initial condition* is the id of the distribution a perturbed genes was initially sampled from. (0) $\mathcal{N}(0.5, 0.1)$, (1) $\mathcal{N}(0.4, 0.05)$, (2) $\mathcal{N}(0.4, 0.1)$ and (3) $\log\mathcal{N}(-2, 1)$. *Perturbation type* refers to how many genes were perturbed (see methods).

Name	Perturbation signal type				Initial condition	Perturbation type
	Fixed	Down	Up	Pulse		
1	0.50	0.17	0.17	0.17	0	multi
2	0.75	0.25	0.00	0.00	0	multi
3	0.75	0.00	0.25	0.00	0	multi
4	0.75	0.00	0.00	0.25	0	multi
5	1.00	0.00	0.00	0.00	0	multi
6	0.50	0.17	0.17	0.17	0	multi
7	0.75	0.25	0.00	0.00	0	multi
8	0.75	0.00	0.25	0.00	0	multi
9	0.75	0.00	0.00	0.25	0	multi
10	1.00	0.00	0.00	0.00	0	multi
11	0.50	0.17	0.17	0.17	0	single
12	0.75	0.25	0.00	0.00	0	single
13	0.75	0.00	0.25	0.00	0	single
14	0.75	0.00	0.00	0.25	0	single
15	1.00	0.00	0.00	0.00	0	single
16	0.50	0.17	0.17	0.17	1	multi
17	0.75	0.25	0.00	0.00	1	multi
18	0.75	0.00	0.25	0.00	1	multi
19	0.75	0.00	0.00	0.25	1	multi
20	1.00	0.00	0.00	0.00	1	multi
21	0.50	0.17	0.17	0.17	2	multi
22	0.75	0.25	0.00	0.00	2	multi
23	0.75	0.00	0.25	0.00	2	multi
24	0.75	0.00	0.00	0.25	2	multi
25	1.00	0.00	0.00	0.00	2	multi
26	0.50	0.17	0.17	0.17	1	single
27	0.75	0.25	0.00	0.00	1	single
28	0.75	0.00	0.25	0.00	1	single
29	0.75	0.00	0.00	0.25	1	single
30	1.00	0.00	0.00	0.00	1	single
31	0.50	0.17	0.17	0.17	2	single
32	0.75	0.25	0.00	0.00	2	single
33	0.75	0.00	0.25	0.00	2	single
34	0.75	0.00	0.00	0.25	2	single
35	1.00	0.00	0.00	0.00	2	single
36	0.50	0.17	0.17	0.17	3	single
37	0.75	0.25	0.00	0.00	3	single
38	0.75	0.00	0.25	0.00	3	single
39	0.75	0.00	0.00	0.25	3	single
40	1.00	0.00	0.00	0.00	3	single
41	0.50	0.17	0.17	0.17	3	multi
42	0.75	0.25	0.00	0.00	3	multi
43	0.75	0.00	0.25	0.00	3	multi
44	0.75	0.00	0.00	0.25	3	multi
45	1.00	0.00	0.00	0.00	3	multi

Table S4. Parametrization of the parameters of the perturbation signal (equation 1) for the four categories fixed signal, decreasing signal, increasing signal and pulsed signal.

Parameter	Fixed	Down	Up	Pulse
h_0	h_{max}	h_{max}	h_{min}	h_{min}
h_1	h_{max}	h_{max}	h_{max}	h_{max}
h_2	h_{max}	h_{min}	h_{max}	$\sim \mathcal{U}(h_{min}, h_{min} + 20)$
t_1	0	0	$\sim \mathcal{U}(0, 1)$	$t_{mid} - t_{center}/2$
t_2	1	$\sim \mathcal{U}(0, 1)$	1	$t_{mid} + t_{center}/2$
b_1	1	$\sim \mathcal{U}(5, 25)$	$\sim \mathcal{U}(5, 25)$	$\sim \mathcal{N}(b_{mean}, 0.25b_{mean}), b_{mean} = 9.1902/t_{width}$
b_2	1	b_1	b_1	$\sim \mathcal{N}(b_{mean}, 0.25b_{mean})$

Table S5. Linear fixed effects models for $F_{\max}$ (BIC -2436.95, R^2 0.674) with predictor variables *A0Init* (initialisation of perturbed gene activation), *Sig* (type of perturbation signals used), *Stdize* (whether input data was standardized per network), *AugRef* (application of randomly sampled reference for $\log_2$ ratio) and *MulPert* (multiple genes perturbed at once). Confidence intervals *CI* for $\alpha = 0.05$.

Name	Coefficient	$\log_{10}$(p-value)	CI lower	CI upper
(Intercept)	0.225	-273.014	0.219	0.231
Sig_fix	0.038	-39.007	0.032	0.043
AugRef_1	0.036	-24.896	0.030	0.043
Stdize_true	0.036	-24.790	0.030	0.043
MulPert_1:AugRef_1	0.027	-7.892	0.018	0.036
Sig_fixup	0.024	-18.660	0.019	0.029
Sig_fixdo	0.018	-10.608	0.013	0.023
Sig_fixpul	0.012	-5.033	0.007	0.017
Stdize_true:MulPert_1	-0.011	-1.830	-0.020	-0.002
MulPert_1	-0.020	-8.980	-0.027	-0.014
Stdize_true:MulPert_1:AugRef_1	-0.029	-4.900	-0.042	-0.016
Stdize_true:AugRef_1	-0.031	-10.392	-0.040	-0.022

Table S6. Linear fixed effects models for $F_{\max}$ (BIC -8580.08, R^2 0.670) with predictor variables *A0Init* (initialisation of perturbed gene activation), *Sig* (type of perturbation signals used), *Stdize* (whether input data was standardized per network), *AugRef* (application of randomly sampled reference for $\log_2$ ratio) and *MulPert* (multiple genes perturbed at once). Confidence intervals *CI* for $\alpha = 0.05$.

Name	Coefficient	$\log_{10}$(p-value)	CI lower	CI upper
(Intercept)	0.248	-Inf	0.242	0.254
Stdize_true	0.040	-19.387	0.032	0.049
AugRef_1:Sig_fixup	0.036	-12.141	0.026	0.046
Stdize_true:A0Init_2:MulPert_1	0.035	-9.699	0.024	0.046
AugRef_1:Sig_fix	0.035	-11.263	0.025	0.045
AugRef_1:Sig_fixpul	0.030	-8.727	0.020	0.040
Sig_fixdo	0.030	-16.517	0.023	0.037
AugRef_1	0.028	-14.184	0.021	0.035
Sig_fix	0.023	-9.960	0.016	0.030
Sig_fixpul	0.010	-2.125	0.003	0.016
A0Init_2	-0.008	-2.537	-0.014	-0.003
Stdize_true:AugRef_1:Sig_fixpul	-0.037	-6.550	-0.051	-0.023
Stdize_true:Sig_fixdo	-0.039	-13.572	-0.048	-0.029
Stdize_true:AugRef_1:Sig_fixup	-0.041	-8.124	-0.055	-0.027
Stdize_true:AugRef_1:Sig_fix	-0.045	-9.351	-0.058	-0.031
Stdize_true:MulPert_1	-0.081	-86.571	-0.089	-0.074