Supplementary information for "Improved normalization in quantum-classical hybrid models for anti-cancer drug response prediction"

Takafumi Ito[1], Artem Lysenko[2,3], Tatsuhiko Tsunoda[1,2,3]

[1] Laboratory for Medical Science Mathematics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan

[2] Laboratory for Medical Science Mathematics, Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan

[3] Mathematical Drug Discovery Platform Project, Program for Drug Discovery and Medical Technology Platforms, RIKEN TRIP Headquarters, Yokohama, Japan

Details of grid search

We determined the optimal hyperparameters by cross-validation. We performed grid searches over the range in Table S1 for the hyperparameters and selected the combination that recorded the highest average AUC. The total number of parameters for the model using the selected hyperparameters is summarized in Table S2. We plotted the prediction performance of each model with the selected hyperparameters by epoch (Figure S1, S2, S3, S4, S5), showing the average of 50 runs of AUC for the validation data obtained from 10 times 5-fold cross-validation. The red dots indicate the points with the highest AUC.

Table S1. Grid search was conducted with this set of hyperparameters.

Hyperparameter	Range
Number of qubits (n_1)	[4, 8]
Number of encoding layers (n_2)	[2, 4]
Number of variational layers (n_3)	[1, 2, 4]
Learning rate	[1e-6, 1e-5, 1e-4]

Table S2. The number of parameters of the models with the selected hyperparameters.

	Cetuximab	Cisplatin	Docetaxel	Erlotinib	Gemcitabine
Classic	7,649,185	7,625,121	7,312,321	7,627,681	7,521,217
Identity	7,651,409	7,625,193	7,310,325	7,629,913	7,521,385
LayerNorm	7,648,185	7,624,121	7,309,253	7,626,685	7,518,137
tanh	7,651,433	7,625,193	7,309,253	7,627,765	7,518,149
Proposed	7,651,417	7,627,353	7,310,325	7,627,753	7,518,149

Table S3. Maximum and minimum values of the output from the classical neural network within the proposed model before training.

	Cetuximab	Cisplatin	Docetaxel	Erlotinib	Gemcitabine
max	0.969	1.051	0.465	1.823	0.489
min	-1.385	-0.635	-0.913	-1.341	-0.962

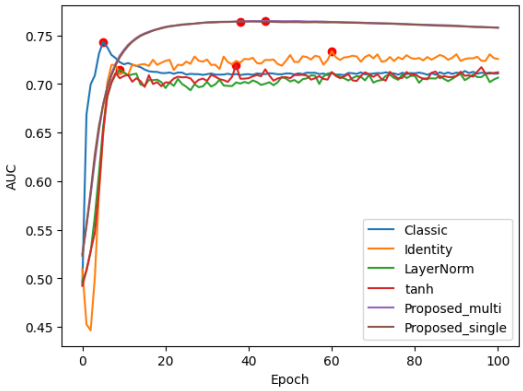


Figure S1. Cetuximab

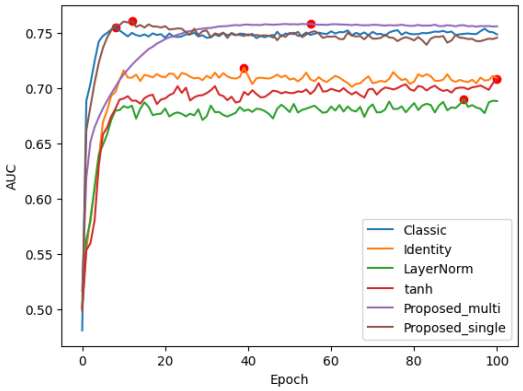


Figure S2. Cisplatin

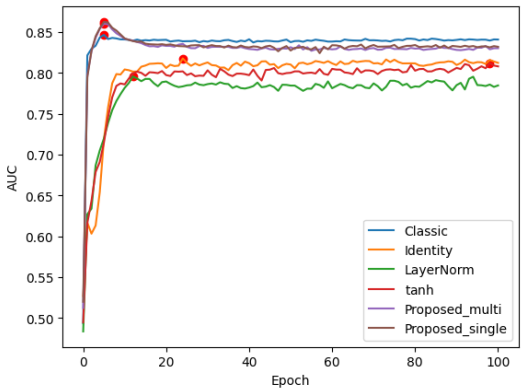


Figure S3. Docetaxel

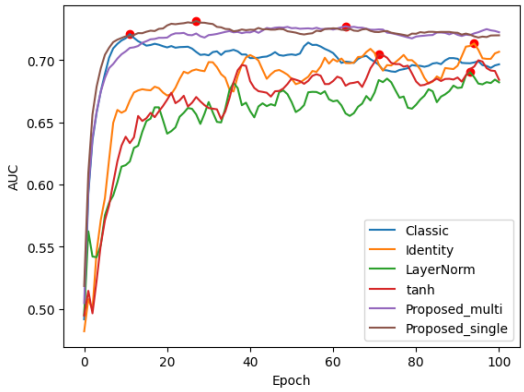


Figure S4. Erlotinib

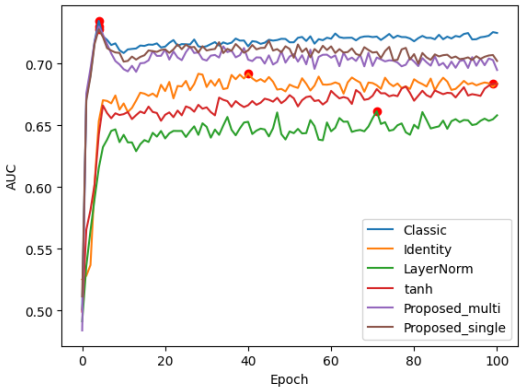


Figure S5. Gemcitabine

Integrated gradient analysis

We conducted enrichment analysis of key genes by Metascape in the “Integrated gradient analysis” section. Results are shown in Figure S6.

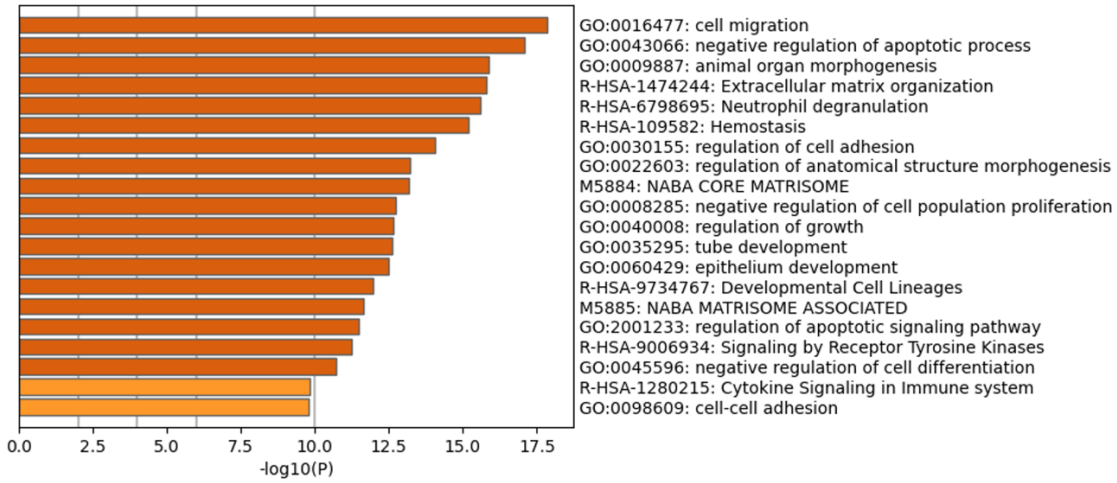


Figure S6. Results of enrichment analysis of key genes by Metascape.

Validation on Quantinuum Reimei H1

We plotted the relationship between shots and AUC to estimate the number of shots that are necessary based on the execution results on Quantinuum Reimei H1 described in the “Validation on the quantum device” section. Since the order of measurements observed in the execution was not recorded, we randomly shuffled the count data for each measurement. We then considered the AUC calculated from the first to the kth measurement as the AUC when we set the number of shots to k. The results are shown in Figure S7, suggesting that 1000 shots are the recommended number of shots, and fewer shots may lead to lower performance. The total execution time for 83 samples with 1,000 shots per sample was 2 minutes 9 seconds from Submitted to Queued, 5 hours 23 minutes 1 second from Queued to Running, and 10 hours 3 minutes 7 seconds from Running to Completed.

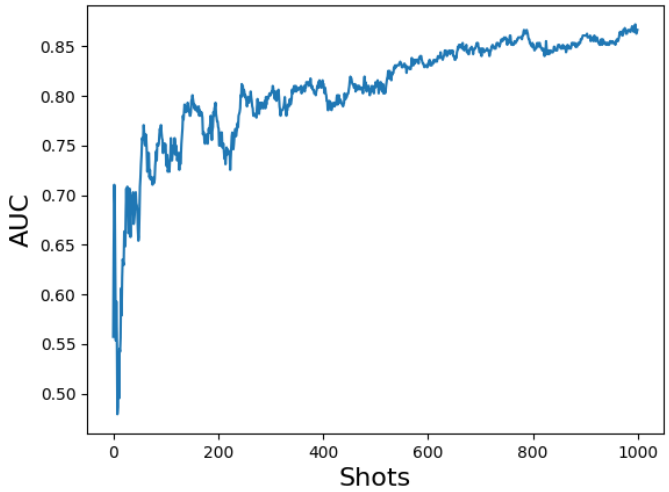


Figure S7. The relationship between shots and AUC