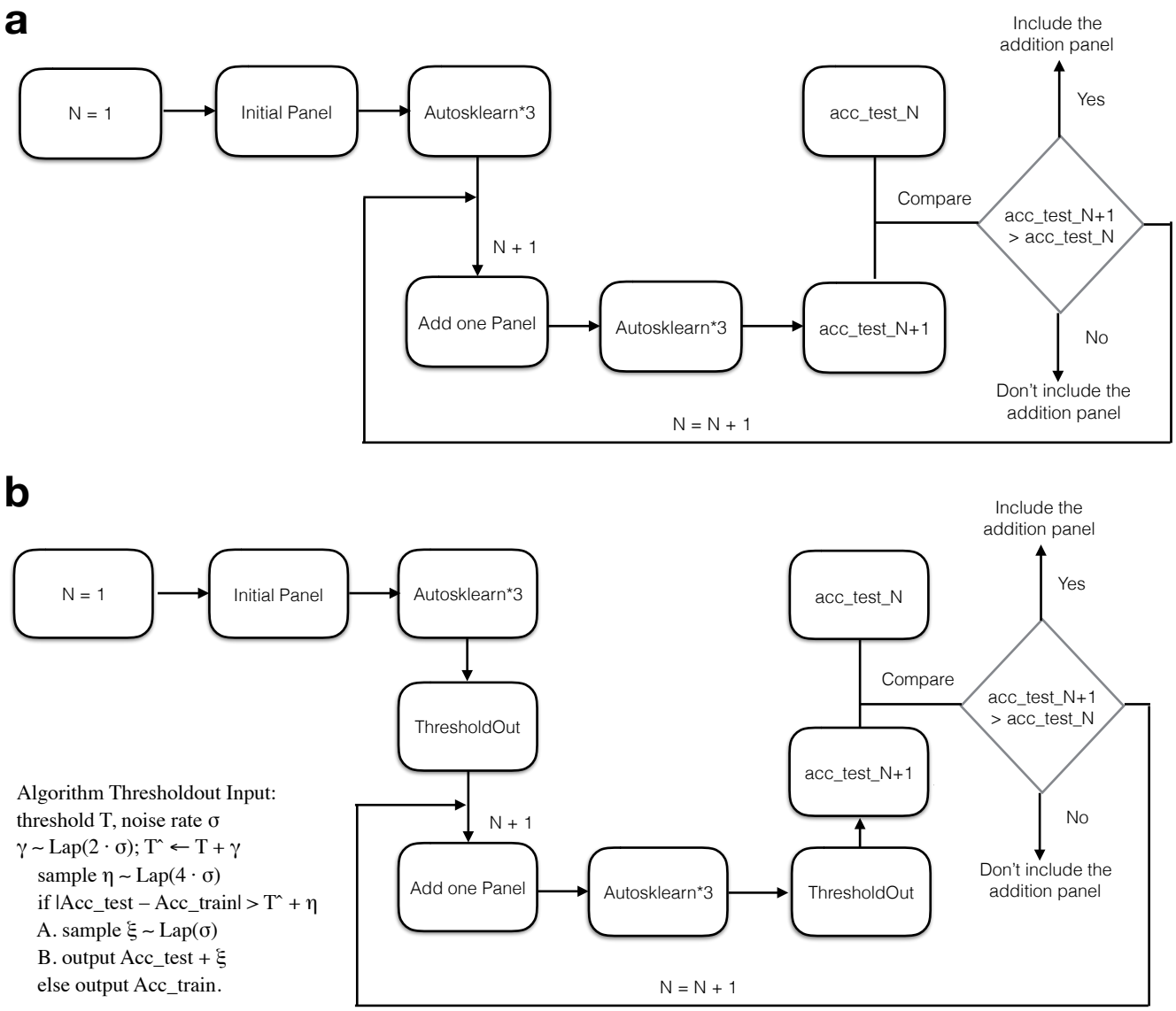
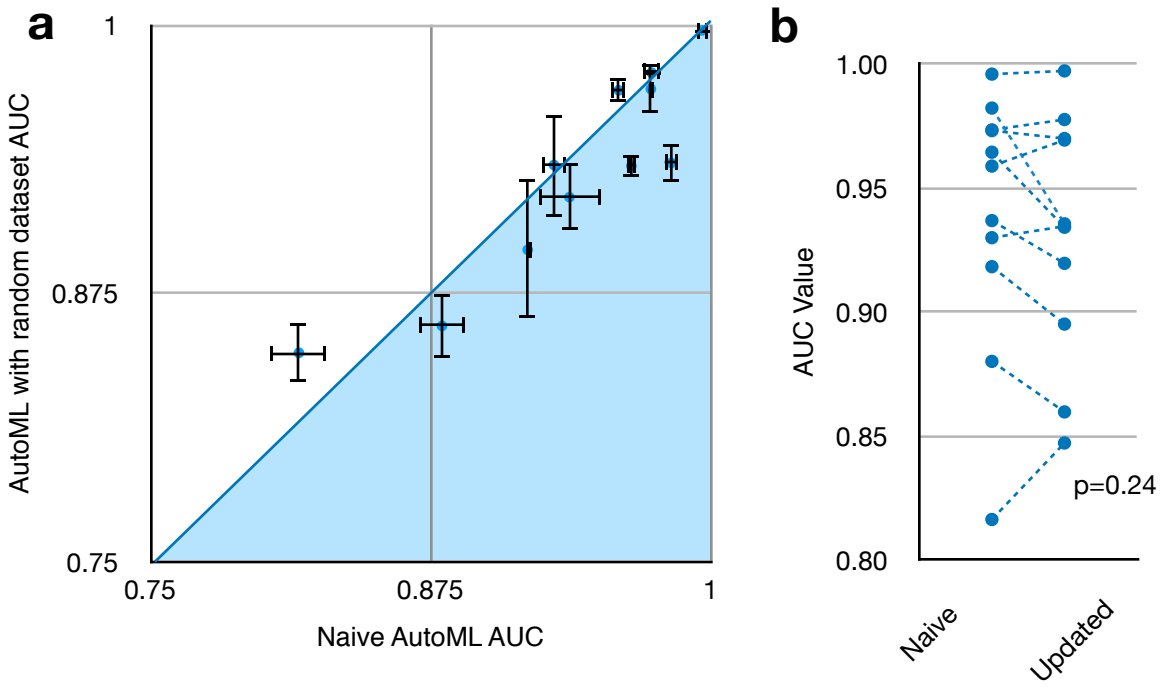
Electronic Supplementary Material (ESI) for Lab on a Chip.
This journal is © The Royal Society of Chemistry 2020

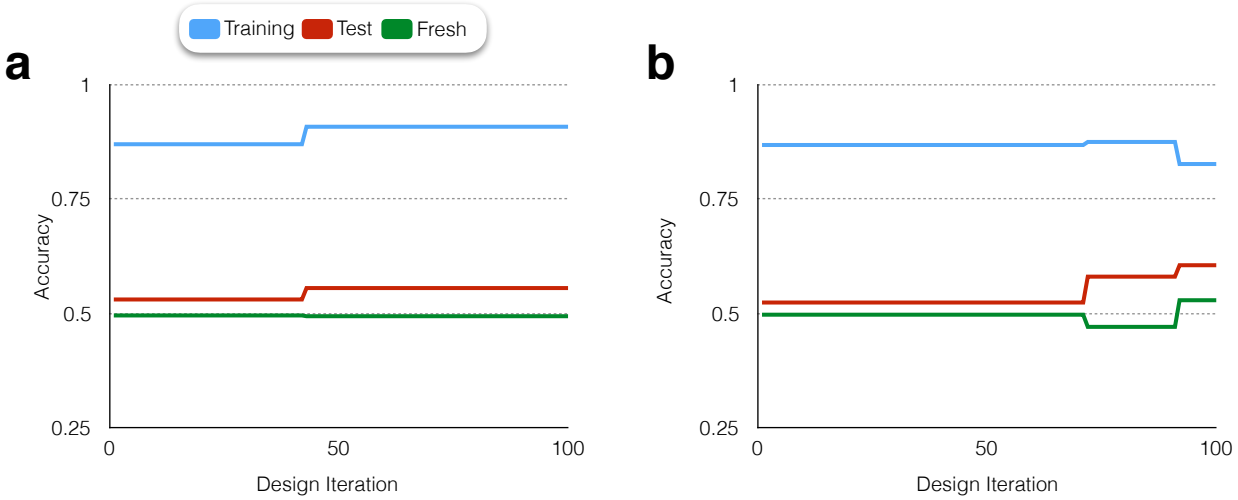
Supplementary Information: A Web-based Automated Machine Learning Platform to Analyze Liquid Biopsy Data



SI Fig.1. Schematic of overfitting simulation experiment. a. We directly performed AutoML training with Standard Reuse (SR) of the test set b. We integrated a Differential Privacy (DP) algorithm - ThresholdOut (TO) into AutoML to provide AutoML with resistance to overfitting. The pseudocode of TO is displayed.



SI Fig.2. Analysis of performance change of AutoML on liquid biopsy datasets from integrating non-liquid-biopsy datasets into AutoML's data repository. a. Comparison between AUC values for 11 datasets by naive AutoML and AutoML updated with 10 non-liquid-biopsy datasets. Error bar represents standard error between the resulting AUCs from 5 individual AutoML training and evaluation for each dataset. b. One-on-one comparison between AUC values generated by updated and naive AutoML, the ΔAUC has a mean of -0.0081 and an averaged propagated standard error of 0.0066.



SI Fig.3. Evaluating the capability of our AutoML to control overfitting with Differential Privacy (DP) algorithm. a.The accuracy of the training set, the test set, and the fresh test set of AutoML implemented with DP, the configuration of DP is with a threshold of 0.1 and a Laplacian noise with a scale of 0.005. b.The accuracy of the training set, the test set, and the fresh test set of AutoML implemented with DP, the configuration of DP is with a threshold of 0.1 and a Laplacian noise with a scale of 0.01.

Data Name	LIterature reported AUC	Automated Machine Learning AUC	Unpaired t-test p value
Ko et.al, 2017 [1]	1.000	0.950±0.007	0.07
Bianchi et.al, 2011 [2]	0.890	0.908±0.005	0.03
Huang et,al, 2016 [3,4]	0.968	0.983±0.005	0.86
LaBreche et.al, 2011 [5]	0.970	0.974±0.002	0.85
LaBreche et.al, 2011 [5]	1.000	0.995±0.002	0.68
Wurdinger et.al, 2015 [6]	0.986	0.969±0.002	0.21
Wurdinger et.al, 2017 [7]	0.910	0.974±0.008	0.14
Ko et.al, 2019 (human) [8]	0.840	0.950±0.009	0.46
Ko et.al, 2019 (murine) [8]	0.800	0.86±0.02	0.04
Clelland et.al, 2013 [9]	N/A	0.934±0.006	0.03
Koth et.al, 2011 [10]	0.951	0.969±0.003	0.37

SI Table.1 Performance comparison between updated AutoML, naive AutoML and literature-adopted customized algorithm. For each dataset, the AUC value generated by updated AutoML was compared with its counterparts generated by naive AutoML and reported by literature. Stand errors were used to show the fluctuation of stochastic prediction results from repeated AutoML trainings. Unpaired T-tests were applied to compare the repeated AutoML performance on the same datasets with the same training-test set partition, and a $p<0.05$ indicated a significant change of AutoML performance after update on its corresponding dataset.

References:

1. J. Ko, N. Bhagwat, S. S. Yee, N. Ortiz, A. Sahmoud, T. Black,N. M. Aiello, L. McKenzie, M. O'Hara, C. Redlingeret al., ACS Nano, 2017,11, 11182–11193.
2. F. Bianchi, F. Nicassio, M. Marzi, E. Belloni, V. Dall'Olio,L. Bernard, G. Pelosi, P. Maisonneuve, G. Veronesi and P.P. Di Fiore, EMBO Molecular Medicine, 2011, 3, 495–503.
3. S. Huang, N. Chong, N. E. Lewis, W. Jia, G. Xie and L. X. Garmire, Genome Medicine, 2016, 8, 34.
4. C. G. A. Network, et al.,Nature, 2012, 490, 61.
5. H. G. LaBreche, J. R. Nevins and E. Huang, BMC Medical Genomics, 2011,4, 61.
6. M. G. Best, N. Sol, I. Kooi, J. Tannous, B. A. Westerman,8 I1–9, F. Rustenburg, P. Schellen, H. Verschueren, E. Post, J. Kosteret al., Cancer Cell, 2015, 28, 666–676.
7. M. G. Best, N. Sol, S. GJG, A. Vancura, M. Muller, A.L. N. Niemeijer, A. V. Fejes, T. K. Fat, A. E. Huis, C. Leurset al., Cancer Cell, 2017,32, 238–252.
8. J. Ko, M. Hemphill, Z. Yang, K. Beard, E. Sewell, J. Shallcross,M. Schweizer, D. K. Sandsmark, R. Diaz-Arrastia, J. Kimet al.,Journal of Neurotrauma, 2019, doi.org/10.1089/neu.2018.6220
9. C. L. Clelland, L. L. Read, L. J. Panek, R. H. Nadrich, C. Bancroft and J. D. Clelland, PLoS One, 2013,8, e69082.
10. L. L. Koth, O. D. Solberg, J. C. Peng, N. R. Bhakta, C. P. Nguyenand P. G. Woodruff, American Journal of Respiratory and Critical Care Medicine, 2011,184, 1153–1163.