MLiB: Interpretable model-based data integration for cancer type classification
Supplementary Material

Zoya van Meel

Table 1: Permutation importances of features which appeared in the 500 most important genes of both GE and CN. ACSM3 was the only feature for which has a positive feature importance in both datasets. Each feature was permuted ten times. The cancer associations column shows a short description of some papers which found evidence linking each feature to cancer. The GE and CN columns show the permutation importance of each feature as the decrease in F_1-score (%). * = Non-cancer research. Unless marked otherwise, all research was done *in vitro*.

Gene	GE	CN	Cancer associations
ACSM3	0.269 ± 0.135	0.059 ± 0.048	Deregulated in hepatocellular carcinoma [1].
ALG12	0.000 ± 0.000	-0.126 ± 0.127	Elevated expression in ovarian cancer [2].
ATP2B2	0.000 ± 0.000	-0.058 ± 0.047	Can be upregulated in breast cancer [3]–[5].
CCL23	0.135 ± 0.165	0.000 ± 0.000	Has pro- and anti-cancer properties w.r.t. to cancer cell interaction [6].
CDH15	-0.233 ± 0.233	0.071 ± 0.058	Expression level serves as a breast cancer prognosis marker [7], [8].
C20orf201	0.000 ± 0.000	0.000 ± 0.000	Cancer testis antigen that could be a useful cancer biomarker [9].
LDLRAD3	0.338 ± 0.391	-0.003 ± 0.010	Biomarker for pancreatic cancer diagnosis [10].
NCLN	0.303 ± 0.101	-0.045 ± 0.054	*Knock-out of *C. elegans* homolog enhances induced neuronal death [11].
NLGN1	0.000 ± 0.000	-0.504 ± 0.331	Upregulation in colorectal cancer [12].
PTPN1	0.000 ± 0.000	0.037 ± 0.037	(*in silico*) expression levels correlated with incidence of esophageal cancer [13].
SF3A2	0.255 ± 0.354	0.000 ± 0.000	Abnormal expression in bladder cancer [14].
SPPL2B	0.034 ± 0.101	-0.019 ± 0.056	Expression levels associated with breast cancer [15].
UBE2E2	0.135 ± 0.165	-0.003 ± 0.010	Upregulation in ovarian cancer [16].

References

[1] R. Gopal, K. Selvarasu, P. P. Pandian, and K. Ganesan, "Integrative transcriptome analysis of liver cancer profiles identifies upstream regulators and clinical significance of ACSM3 gene expression," *Cellular oncology*, vol. 40, pp. 219–233, 2017.

[2] X. Cui, X. Pei, H. Wang, *et al.*, "ALG3 promotes peritoneal metastasis of ovarian cancer through increasing interaction of α1, 3-mannosylated uPAR and ADAM8," *Cells*, vol. 11, no. 19, p. 3141, 2022.

[3] W. J. Lee, S. J. Roberts-Thomson, N. A. Holman, F. J. May, G. M. Lehrbach, and G. R. Monteith, "Expression of plasma membrane calcium pump isoform mrnas in breast cancer cell lines," *Cellular signalling*, vol. 14, no. 12, pp. 1015–1022, 2002.

[4] K. Varga, A. Hollósi, K. Pászty, *et al.*, "Expression of calcium pumps is differentially regulated by histone deacetylase inhibitors and estrogen receptor alpha in breast cancer cells," *BMC cancer*, vol. 18, pp. 1–17, 2018.

[5] W. J. Lee, S. J. Roberts-Thomson, and G. R. Monteith, "Plasma membrane calcium-ATPase 2 and 4 in human breast cancer cell lines," *Biochemical and biophysical research communications*, vol. 337, no. 3, pp. 779–783, 2005.

[6] J. Korbecki, K. Kojder, D. Simińska, *et al.*, "CC chemokines in a tumor: A review of pro-cancer and anti-cancer properties of the ligands of receptors CCR1, CCR2, CCR3, and CCR4," *International journal of molecular sciences*, vol. 21, no. 21, p. 8412, 2020.

[7] G. Kremmidiotis, E. Baker, J. Crawford, H. J. Eyre, J. Nahmias, and D. F. Callen, "Localization of human cadherin genes to chromosome regions exhibiting cancer-related loss of heterozygosity," *Genomics*, vol. 49, no. 3, pp. 467–471, 1998.

[8] M. Xu, C. Liu, L. Pu, *et al.*, "Systemic analysis of the expression levels and prognosis of breast cancer-related cadherins," *Experimental Biology and Medicine*, vol. 246, no. 15, pp. 1706–1720, 2021.

[9] M. H. Almutairi, "Identification of novel human cancer-testis-antigen genes in cancers," Ph.D. dissertation, Bangor University (United Kingdom), 2014.

[10] F. Yang, D.-Y. Liu, J.-T. Guo, *et al.*, "Circular RNA circ-LDLRAD3 as a biomarker in diagnosis of pancreatic cancer," *World journal of gastroenterology*, vol. 23, no. 47, p. 8345, 2017.

[11] S. Kamat, S. Yeola, W. Zhang, L. Bianchi, and M. Driscoll, "Nra-2, a nicalin homolog, regulates neuronal death by controlling surface localization of toxic caenorhabditis elegans deg/enac channels," *Journal of Biological Chemistry*, vol. 289, no. 17, pp. 11 916–11 926, 2014.

[12] Q. Yu, X. Wang, Y. Yang, *et al.*, "Upregulated NLGN1 predicts poor survival in colorectal cancer," *BMC cancer*, vol. 21, pp. 1–11, 2021.

[13] J. Chen, X. Zhao, Y. Yuan, and J.-J. Jing, "The expression patterns and the diagnostic/prognostic roles of PTPN family members in digestive tract cancers," *Cancer Cell International*, vol. 20, pp. 1–15, 2020.

[14] K.-L. Liu, Y.-W. Yin, B.-S. Lu, *et al.*, "E2F6/KDM5C promotes SF3A3 expression and bladder cancer progression through a specific hypomethylated DNA promoter," *Cancer Cell International*, vol. 22, no. 1, p. 109, 2022.

[15] Y. Cho, "Studies on the mechanism of regulated intramembrane proteolysis of Prss14 and its role in cancer metastasis," Ph.D. dissertation, 서울대학교 대학원, 2019.

[16] X. Hong, N. Ma, D. Li, *et al.*, "UBE2E2 enhances Snail-mediated epithelial-mesenchymal transition and Nrf2-mediated antioxidant activity in ovarian cancer," *Cell Death & Disease*, vol. 14, no. 2, p. 100, 2023.