



CORRESPONDENCE

False discovery rate control in cancer biomarker selection



We read with great interest the manuscript “LPCAT1 functions as a novel prognostic molecular marker in hepatocellular carcinoma” by Zhang et al in a recent issue of *Genes & Diseases*.¹ The authors conducted bioinformatics analyses using high throughput RNA sequencing data from TCGA to demonstrate that LPCAT1 is a novel and effective prognostic marker for hepatocellular carcinoma. We appreciate the contributions of the authors on the subject, nonetheless, we have some concerns that should be clarified in the following issues.

The standard *P*-value was invented for testing individual hypotheses. There is an obvious problem when analyzing gene expression data collected via sequencing of multiple genomes, as this usually involves testing from several thousands to tens of thousands of hypotheses simultaneously. In genome sequencing studies most researchers are keenly aware of the potentially high rate of false positives and the need to control it. One key statistical shift is the move away from the well-known *P*-value to false discovery rate (FDR).^{2,3} The FDR of a test is defined as the

expected proportion of false positives among the declared significant results.^{3–5} Because of this directly useful interpretation, FDR is a more convenient scale to work on instead of the *P*-value scale. However, in Zhang’s report, the multiple test correction was not applied to the *P* values shown in Figure 2A and 7 and stated in the text. This seems to be required since the authors tested the association between each gene and outcome individually.

The correlation between gene expression and overall survival in the same hepatocellular carcinoma RNA sequencing data from TCGA were re-analyzed by the Genomics Analysis and Visualization Platform (<http://r2.amc.nl>) and the results were corrected for multiple gene testing by FDR. The potential LPCAT1-related tumor genes reported by the author and their adjusted *P*-value were provided in Table 1. It showed that the expression of several genes (CCNB2, CENPF, and UBE2C) had no associations with overall survival, which is different from Figure 7 in the report.

Peer review under responsibility of Chongqing Medical University.

<https://doi.org/10.1016/j.gendis.2022.12.010>

2352-3042/© 2023 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Table 1 The association of LPCAT1-related tumor genes with overall survival. Results were corrected for multiple gene testing by false discovery rate.

No.	Gene	Probeset	Adjusted <i>P</i> -value
1	<i>CDC20</i>	CDC20_991	0.001464779
2	<i>CDCA8</i>	CDCA8_55143	0.005409055
3	<i>LPCAT1</i>	LPCAT1_79888	0.00569335
4	<i>TPX2</i>	TPX2_22974	0.00691434
5	<i>DLGAP5</i>	DLGAP5_9787	0.008015183
6	<i>CCNB1</i>	CCNB1_891	0.008149167
7	<i>MAD2L1</i>	MAD2L1_4085	0.009098616
8	<i>KIF4A</i>	KIF4A_24137	0.00910015
9	<i>NUF2</i>	NUF2_83540	0.009505066
10	<i>CENPA</i>	CENPA_1058	0.011305291
11	<i>KIF11</i>	KIF11_3832	0.011942813
12	<i>KIF20A</i>	KIF20A_10112	0.012491643
13	<i>BIRC5</i>	BIRC5_332	0.013002953
14	<i>KIF2C</i>	KIF2C_11004	0.013481529
15	<i>CDK1</i>	CDK1_983	0.015119904
16	<i>TTK</i>	TTK_7272	0.018113202
17	<i>PLK1</i>	PLK1_5347	0.018869319
18	<i>BUB1</i>	BUB1_699	0.023996996
19	<i>BUB1B</i>	BUB1B_701	0.025563704
20	<i>RRM2</i>	RRM2_6241	0.025577274
21	<i>NCAPG</i>	NCAPG_64151	0.029011476
22	<i>TOP2A</i>	TOP2A_7153	0.029439732
23	<i>NDC80</i>	NDC80_10403	0.029498546
24	<i>RACGAP1</i>	RACGAP1_29127	0.030345955
25	<i>KIF18A</i>	KIF18A_81930	0.035375811
26	<i>CEP55</i>	CEP55_55165	0.036683405
27	<i>CENPE</i>	CENPE_1062	0.041494437
28	<i>AURKB</i>	AURKB_9212	0.047951168
29	<i>CCNB2</i>	CCNB2_9133	not significant
30	<i>CENPF</i>	CENPF_1063	not significant
31	<i>UBE2C</i>	UBE2C_11065	not significant

Conflict of interests

The author declares no potential conflict of interests.

References

- Zhang H, Xu K, Xiang Q, et al. LPCAT1 functions as a novel prognostic molecular marker in hepatocellular carcinoma. *Genes Dis.* 2022;9(1):151–164.
- Pawitan Y, Michiels S, Koscielny S, et al. False discovery rate, sensitivity and sample size for microarray studies. *Bioinformatics.* 2005;21(13):3017–3024.
- Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A.* 2003;100(16):9440–9445.
- Shen A, Fu H, He K, et al. False discovery rate control in cancer biomarker selection using knockoffs. *Cancers.* 2019;11(6):744.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol.* 1995;57(1):289–300.

Zhaoming Li

Department of Oncology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450052, China
E-mail address: fcclizm@zzu.edu.cn

17 October 2022