Ulrich Mahlknecht, MD, PhD

Editor-in-Chief

Clinical Epigenetics

December 6th, 2014

Dear Dr. Mahlknecht,

Thank you for your letter on Dec 6th, 2014 regarding our manuscript entitled “Identification and validation of the methylation biomarkers of non-small cell lung cancer (NSCLC)”. Our appreciation also goes to the reviewers for their helpful comments. We have revised the manuscript following the reviewer’s comments and your instructions.

Enclosed please find the revised version of the manuscript along with a point by point description of our responses to the reviewer’s comments. We have used coloured text for all changes made in the revised manuscript.

We hope that the manuscript is now acceptable for publication in Clinical Epigenetics. Thank you again for your information and editorial assistance.

Sincerely yours,

Jiucun Wang, Ph.D.

Professor

School of Life Sciences

Fudan University Jiangwan Campus

2005 Songhu Road, Shanghai 200438

People’s Republic of China.

Tel: (+) 86-21-51630606

Fax: (+) 86-21-51630607

E-mail: [jcwang@fudan.edu.cn](mailto:jcwang@fudan.edu.cn)

**Responses to Reviewer 1’s comments**

**Comment:** I would suggest that the authors could discuss the Multi-cellular issue a bit in the manuscript, not only mention this in the response letter to reviewers.

**Response:** Thanks for your suggestion. We have added the discussion about the multi-cellular issue into the section of Discussion in the revised manuscript with red marked text.

**Responses to Reviewer 2’s comments**

**Comment:**

I believe there must be a mistake either in the standard error calculation, or in the data reported in Table 2. For the SLC5A8 example, following the steps outlined in http://www.bmj.com/content/343/bmj.d2304, the standard error should be 0.658. Given log10(OR) of 3.80, the z score is 5.775 and the corresponding p-value is 9E-7. Using log10 or logE doesn't matter here as long as we've been consistent (the SE also reported in log10), because the log bases will cancel each other in the z score.

I was not able to replicate the authors' calculation to get the standard error of 1.1515 using their equations 2-4 (in the cover letter) and data in Table 2 [log10OR=3.80, 95% CI for log10OR=(2.51, 5.09)].

I'd ask the authors to review their calculations and if possible, provide the output of this SLC5A8 analysis. Assuming the authors obtained the results using the equation 1 in their cover letter, the output of 'summary(glm)' in R would give us all the details necessary to figure out what went wrong.

**Response:**

Thanks so much for your careful checking! We checked Table 2 again and re-analyzed the data, and found there was something wrong with the footnote of Table 2. After the modification of the footnote of Table 2, all the calculations met your deduction. We thank you for your help again. We have changed the footnote of the table and incorporated the corresponding description in the section of the Result in the revised version with red colored text.

Actually, P-valuea is the bonferroni adjusted P-value which were based on paired *t*-test compared the intensity of the methylation signals between case and control. log10(OR) and P-valueb represent log-transformed Odds ratio and P-value based on logistic regression adjusted by gender, age and smoking status.

As we have mentioned in our previous response, the beta and P-value were calculated by the logistic regression in the R code.

glm<- glm(y~gender+age+smoking+x,data,family=binomial(logit))) (1)

beta and the standard error (se) were estimated by the above R code. Be careful, the value in the column 4 is based on log10 rather than logE. We can calculate OR and the 95% CI of the OR by the following functions.

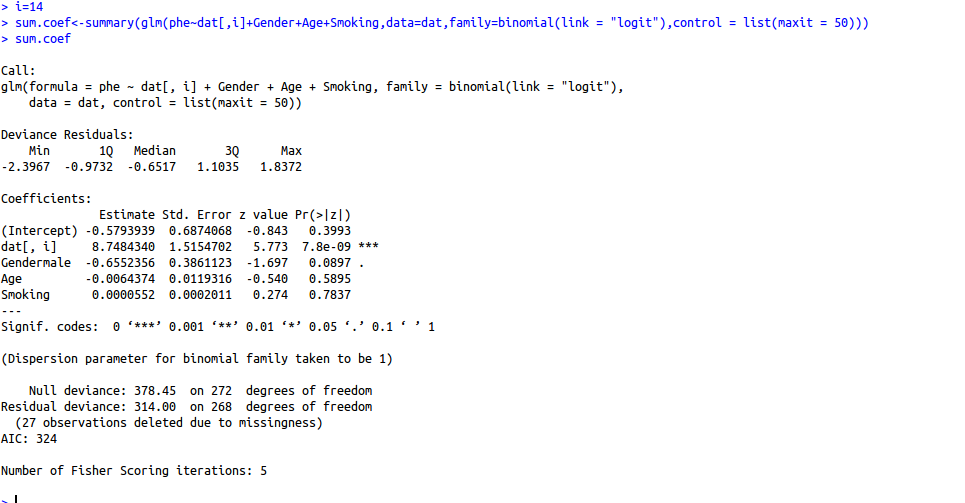
OR <- exp(beta) (2)

Up <- exp(beta+1.96\*se) (3)

Low <- exp(beta-1.96\*se) (4)

You can find our analysis process as the following:

SLC5A8 is the 14th column of the dataset.



You can find that the standard error is 1.515, the coefficient of the SLC5A8 is 8.748, the Z score is 5.773 and the corresponding P-value is 7.80E-9 (i.e. 2\*(1-pnorm(5.773))).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2. Differential Methylation in NSCLCs | | | | | | | | |
|  | AMP  (NSCLC) | AMP  (Control) | P-valuea | log10(OR) (95%CI) | P-valueb | Sen | Spe | AUC |
| *AGTR1* | 12.88% | 4.48% | 1.06E-07 | 3.49 (2.08, 4.91) | 1.30E-06 | 59.73% | 79.59% | 0.71 |
| *GALR1* | 18.31% | 2.91% | 6.58E-09 | 2.56 (1.5, 3.63) | 2.30E-06 | 46.98% | 85.03% | 0.67 |
| *NTSR1* | 9.37% | 0.56% | 1.09E-09 | 9.02 (5.48, 12.55) | 5.90E-07 | 44.30% | 94.56% | 0.70 |
| *SLC5A8* | 25.59% | 11.66% | 4.77E-12 | 3.80 (2.51, 5.09) | 7.80E-09 | 52.35% | 88.44% | 0.67 |
| *ZMYND10* | 6.95% | 12.82% | 1.08E-07 | -4.61 (-6.27, -2.95) | 5.20E-08 | 73.15% | 92.52% | 0.80 |
| *LINE-1* | 72.10% | 76.76% | 2.39E-12 | -10.3 (-13.5, -7.2) | 1.80E-10 | - | - | - |
| Reference | 1.78% | 1.83% | 2.85E-01 | -19.37 (-45.35, 6.62) | 0.14 | - | - | - |

Differential methylation analysis was conducted between 150 NSCLC and adjacent normal tissues. AMP: Average methylation percentage. P-valuea is the bonferroni adjusted P-value which were based on paired *t*-test compared the intensity of the methylation signals between case and control. log10(OR) and P-valueb represent log-transformed Odds ratio and p-value based on logistic regression adjusted by gender, age and smoking status. Reference site was a C site that was not in CpG site, therefore, no or low methylated signal would be detected and non-significant association should be detected between cancer and normal tissues. Sensitivity, specificity and AUC were calculated by logistic regression prediction model without adjustment for gender, age and smoking status.